

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 23:03:01 ; Search time 81 Seconds

(without alignments)  
981.904 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038

Sequence: 1 MLPKRMKLLPLVLSQMAIL.....EORNLLDLSTWVPEQIH 386

Scoring table: BLAST62

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.21.\*  
1: sp.\_archaea.\*  
2: sp.\_bacteria.\*  
3: sp.\_fungi.\*  
4: sp.\_human.\*  
5: sp.\_invertebrate.\*  
6: sp.\_mammal.\*  
7: sp.\_mhc.\*  
8: sp.\_organelle.\*  
9: sp.\_phage.\*  
10: sp.\_plant.\*  
11: sp.\_rodent.\*  
12: sp.\_virus.\*  
13: sp.\_vertebrate.\*  
14: sp.\_unclassified.\*  
15: sp.\_virus.\*  
16: sp.\_bacteria.\*  
17: sp.\_archae.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	4	Q9Y5R3
2	1515.5	74.4	388	11	Q9RL11
3	1513.5	74.3	388	11	Q9WUE5
4	1019	50.0	395	4	Q9G2X3
5	1017.5	49.9	395	11	Q9GUP4
6	1008	49.5	411	4	Q9GZS9
7	627	30.8	484	11	Q9EP78
8	624	30.6	484	11	Q9NP80
9	606.5	29.8	486	4	Q7367
10	605.5	29.7	486	4	Q9NS84
11	598.5	29.4	530	11	Q9B276
12	591.5	29.0	530	4	Q9Y6F2
13	587.5	28.8	484	4	Q9UCD5
14	587.5	28.8	531	4	Q9Y4C5
15	559	27.4	411	11	Q9BQC0
16	549	26.9	411	4	Q43516

ID	Q9Y5R3	PRELIMINARY:	PRT:	386 AA.
AC	Q9Y5R3;			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand			
DE	sulfotransferase GST-3).			
GN	GST3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid-9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=TONSIL;			
RX	MEDLINE=9264336; PubMed=10330415;			
RA	Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,			
RA	Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;			
RT	"Sulfotransferases of two specificities function in the reconstitution			
RT	of high endothelial cell ligands for L-selectin.";			
RL	J. Cell Biol. 145:899-910(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=TONSIL;			
RA	Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddie N.R.,			
RA	Rosen S.D.;			
RT	"Chromosomal Localization and Genomic Organization for the			
RT	Galactose/4-Acetylglucosamine/N-Acetylglucosamine 6-O-			
RT	Sulfotransferase Gene Family.";			
RL	Glycobiology 0:0-0(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=2132592; PubMed=11439191;			
RA	Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Eljies L.G.,			
RA	Rabuka D., Hindsaugh O., Martin J.D., Lowe J.B., Fakuda M.;			
RT	"Novel sulfated lymphocyte homing receptors and their control by a			
RT	core1 extension beta1,3-N-acetylglucosaminyltransferase.";			

## ALIGNMENTS

ID	Q9Y5R3	PRELIMINARY:	PRT:	386 AA.	
17	504.5	24.8	472	11	Q88199
18	482	23.7	479	4	Q75099
19	481.5	23.6	474	11	Q8Q212
20	423.5	20.8	441	13	Q83403
21	264	13.0	486	5	Q95T19
22	210	10.3	363	5	Q9VWC4
23	203	10.0	114	11	Q9D0K5
24	186.5	9.2	183	5	Q9VWC2
25	175	8.6	119	6	Q95T18
26	168	8.7	307	16	Q82V64
27	113.5	5.7	303	16	Q83UB6
28	111	5.4	655	6	Q9B6X0
29	107	5.3	1586	4	Q9P2D3
30	100.5	4.9	747	4	Q960U2
31	100.5	4.9	775	4	Q960U1
32	100	4.9	593	10	Q8VZE0
33	96.5	4.7	743	4	Q9H0K2
34	96	4.7	2454	3	Q9UVF2
35	96	4.7	2454	3	Q9UV56
36	92.5	4.5	2354	5	Q24135
37	92.5	4.5	2354	5	Q9VXG8
38	91.5	4.5	338	11	Q354D0
39	91.5	4.5	338	11	Q21V03
40	91.5	4.5	4631	4	Q8WXY2
41	91.5	4.5	5314	4	Q8WXY1
42	91.5	4.5	5938	4	Q96PK2
43	91	4.5	294	11	Q9CZJ0
44	91	4.5	1847	17	Q58817
45	90	4.4	573	5	Q9GPG0

RL Cell 105:957-969(2001).  
 DR EMBL, AF131235; AAD33015.1; -  
 DR EMBL, AF280088; AAG48246.1; -  
 DR EMBL, AF149783; AAK48417.1; -  
 DR InterPro: IPR001092; HLH\_basic.  
 DR PROSITE: PS00038; Helix\_LOOP\_Helix; UNKNOWN\_1.  
 KW Lectin; Selectin; Transferrase.  
 SQ SQUONCE 386 AA; 45133 MW; 0C3BBA02241743A CRC64;

Query Match 100.0%; Score 2038; DB 4; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-179;  
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPKKKLLFLVSGMATALFHHYSHNISLSKKAQPERHNVLYLSMRSGSSFGQ 60  
 DB 1 MLLPKKKLLFLVSGMATALFPHMYSHNISLSKKAQPERHNVLYLSMRSGSSFGQ 60  
 QY 61 LFGQHPDVFYLMPEPAWVMTFKOSTAMMLHMAVRDLRAVELCDMSVDAVMEPPRRQ 120  
 DB 61 LFGQHPDVFYLMPEPAWVMTFKOSTAMMLHMAVRDLRAVELCDMSVDAVMEPPRRQ 120  
 QY 121 SSLEFOWNSRSLCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSYSHVYLKEV 180  
 DB 121 SSLEFOWNSRSLCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSYSHVYLKEV 180  
 QY 181 FFMLOSILYPLKDPSLNLHVLVDRPRAVRSRERTKGDLMDSRIYMGQHEQKLKED 240  
 DB 181 FFMLOSILYPLKDPSLNLHVLVDRPRAVRSRERTKGDLMDSRIYMGQHEQKLKED 240  
 QY 241 QPYVWGVICQSOLEIYKTQSLPKAJOERYLLVREYDLARAVAOISRMVEFGLEFL 300  
 DB 241 QPYVWGVICQSOLEIYKTQSLPKAJOERYLLVREYDLARAVAOISRMVEFGLEFL 300  
 QY 301 HLQTVWNIIRGKMGDHAFTNARALNVSQAMRWSLPEYKYSRLQKACGDAMNLGYR 360  
 DB 301 HLQTVWNIIRGKMGDHAFTNARALNVSQAMRWSLPEYKYSRLQKACGDAMNLGYR 360  
 QY 361 HVRSQEQRNLLDLSTWTPVPOIH 386  
 DB 361 HVRSQEQRNLLDLSTWTPVPOIH 386

## RESULT 2

Q9R1I1 PRELIMINARY: PRT; 388 AA.

AC Q9R1I1;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE L-selectin ligand sulfotransferase.

GN CHST4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99361934; PubMed=10435581;  
 RA Hirooka N., Petrjnlak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,  
 RA Iwawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.,  
 RT "A novel, high endothelial venule-specific sulfotransferase expresses  
 RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34,"  
 RT Immunoty 11:79-89(1999).  
 DR EMBL: AF109155; AAD45579.1; -  
 DR MGD: MGI1349479; Chst4.  
 KW Lectin; Selectin; Transferrase.  
 SQ SQUONCE 388 AA; 44636 MW; 6D5371AFB6884AEE CRC64;

Query Match 74.4%; Score 1515.5; DB 11; Length 388;  
 Best Local Similarity 72.8%; Pred. No. 2e-131;  
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;  
 QY 1 MLLPKKKLLFLVSGMATALFHHYSHNISLSKKAQPERHNVLYLSMRSGSSFGV 59

DB 1 MLLPKKKLLFLVSGQIVVALFTIHMSVHR--HLISQRESRPVHLYLSMRSGSSFGV 58  
 QY 60 QLFQHPDVFYLMPEPAWVMTFKOSTAMMLHMAVRDLRAVELCDMSVDAVMEPPRR 119  
 DB 59 QLFQHPDVFYLMPEPAWVMTFKOSTAMMLHMAVRDLRAVELCDMSVDAVMEPPRR 118  
 QY 120 QSLFOWNSRSLCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSYSHVYLKEV 179  
 DB 119 QSLFOWNSRSLCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSYSHVYLKEV 178  
 QY 180 FFMLOSILYPLKDPSLNLHVLVDRPRAVRSRERTKGDLMDSRIYMGQHEQKLKED 239  
 DB 179 FFMLOSILYPLKDPSLNLHVLVDRPRAVRSRERTKGDLMDSRIYMGQHEQKLKED 238  
 QY 240 DQPYWGVICQSOLEIYKTQSLPKAJOERYLLVREYDLARAVAOISRMVEFGLEFL 299  
 DB 239 DQPYWGVICQSOLEIYKTQSLPKAJOERYLLVREYDLARAVAOISRMVEFGLEFL 298  
 QY 300 HLQTVWNIIRGKMGDHAFTNARALNVSQAMRWSLPEYKYSRLQKACGDAMNLGYR 359  
 DB 299 HLQTVWNIIRGKMGDHAFTNARALNVSQAMRWSLPEYKYSRLQKACGDAMNLGYR 358  
 QY 360 HVRSQEQRNLLDLSTWTPVPOIH 385  
 DB 359 HVRSQEQRNLLDLSTWTPVPOIH 384

## RESULT 3

Q9WUE5 PRELIMINARY: PRT; 388 AA.

AC Q9WUE5;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin  
 DE 6/veratan) sulfotransferase 4).  
 GN CHST4.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=TONGUE;  
 RX MEDLINE=99264336; PubMed=10330415;  
 RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zhu F.-R.,  
 RA Huang C.-C., Kanagaki R., Rosen S.D., Hemmerich S.,  
 RT "Sulfotransferases of two specificities function in the reconstitution  
 RT of high endothelial cell ligands for L-selectin,"  
 RT J Cell Biol. 145:899-910(1999).  
 RL [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Yamana K.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staibli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Bairdrelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaudo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gerbaldi M.,  
 RA Gustinchich S., Hall D., Hoffmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki T.;



RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AF131236; AAD3016.1; -  
 DR EMBL: AK009113; BAB26078.1; -  
 DR MGD: MGI:1349479; Chst4.  
 KW Transferase.  
 SO SEQUENCE 388 AA; 44694 MW; D1E9D796DF8574D CRC64;

Query Match 74.38; Score 1513.5; DB 11; Length 388;  
 Best Local Similarity 72.8%; Pred. No. 3e-131;  
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MLLPKMKLLFLVSQMAILLALFFHMYSHNITSLSMKAOPERMHVIVLSSWRSGSSFFVQ 59  
 1 MMLLKKGLMLFLSQVIVLALFTHMSYHR--HLSDREBSRPVHVIVLSSWRSGSSFFVQ 58  
 Db 1 MMLLKKGLMLFLSQVIVLALFTHMSYHR--HLSDREBSRPVHVIVLSSWRSGSSFFVQ 58  
 QY 60 QLFQHPDVFYLMERPAMHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRR 119  
 59 QLFQHPDVFYLMERPAMHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRR 118  
 Db 59 QLFQHPDVFYLMERPAMHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRR 118  
 QY 120 QSLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVERKACRSYHVYLKEV 179  
 119 QSLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVERKACRSYHVYLKEV 178  
 Db 119 QSLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVERKACRSYHVYLKEV 178  
 QY 180 REFNIQSLYPLIKDPSLNLHIVLVRDPRAVFRSRETKGDLMTDSRIVWQHEQKLKE 239  
 179 REFNIQSLYPLIKDPSLNLHIVLVRDPRAVFRSRETKGDLMTDSRIVWQHEQKLKE 238  
 Db 179 REFNIQSLYPLIKDPSLNLHIVLVRDPRAVFRSRETKGDLMTDSRIVWQHEQKLKE 238  
 QY 240 DQPIYVMOVYCQSOLEIKYK--TIQSLPKALOERYLVREEDLARAPVQTSRYEFGLEFL 299  
 239 DQPIYVMOVYCQSOLEIKYK--TIQSLPKALOERYLVREEDLARAPVQTSRYEFGLEFL 298  
 Db 239 DQPIYVMOVYCQSOLEIKYK--TIQSLPKALOERYLVREEDLARAPVQTSRYEFGLEFL 298  
 QY 300 PHLQWVHNITRGKMGNDHAFHTNARDALNVSQAMRSLPEYKYSRLQKAGDANLLG 359  
 299 PHLQWVHNITRGKMGNDHAFHTNARDALNVSQAMRSLPEYKYSRLQKAGDANLLG 358  
 Db 299 PHLQWVHNITRGKMGNDHAFHTNARDALNVSQAMRSLPEYKYSRLQKAGDANLLG 358  
 QY 360 RHVRSEDEQRNLDDLSTWVPEQI 385  
 359 RHVRSEDEQRNLDDLSTWVPEQI 384  
 Db 359 RHVRSEDEQRNLDDLSTWVPEQI 384

## RESULT 4

Q9GZX3 PRELIMINARY; PRT; 395 AA.

ID Q9GZX3  
 AC Q9GZX3  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-acetylglucosamine-6-O-sulfotransferase).  
 GN GST4BETA OR CHST6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R., Rosen S.D.;  
 RT "Chromosomal Localization and Genomic Organization for the Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-sulfotransferase Gene Family."  
 RT Glycobiology 0:0-0(2001).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RA MBDLINE=20472330; PubMed=11017086;  
 RA Akama T.O., Nishide K., Nakayama J., Matanabe H., Fujiwara T., Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., Ozaki K., Kinoshita S., Shimomura Y., Taniguchi A., Fukuda M.N.;  
 RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene."  
 RT Nat. Genet. 26:237-241(2000).  
 DR EMBL: AF280086; AAG48244.1; -

DR EMBL: AF219990; AAG26325.1; -  
 DR EMBL: AF219991; AAG26327.1; -  
 KW Transferase.  
 SO SEQUENCE 395 AA; 44098 MW; 433CA60248A4BF67 CRC64;

Query Match 50.08; Score 1019; DB 4; Length 395;  
 Best Local Similarity 54.5%; Pred. No. 1.2e-85;  
 Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LPRKMKLLFLVSQMAILLALFFHMYSHNITSLSMKAOPERMHVIVLSSWRSGSSFFVQ 61  
 14 LLAQTFLLFLVSRP-----GPSSPAGGEARVHVIVLSSWRSGSSFFVQ 59  
 Db 14 LLAQTFLLFLVSRP-----GPSSPAGGEARVHVIVLSSWRSGSSFFVQ 59  
 QY 62 FQGHDPVFYLMERPAMHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRR 121  
 60 FQGHDPVFYLMERPAMHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRR 118  
 Db 60 FQGHDPVFYLMERPAMHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRR 118  
 QY 122 SLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVERKACRSYHVYLKEV 181  
 119 SLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVERKACRSYHVYLKEV 178  
 Db 119 SLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVERKACRSYHVYLKEV 178  
 QY 182 FNIQSLYPLIKDPSLNLHIVLVRDPRAVFRSRETKGDLMTDSRIVWQHEQKLKE 241  
 179 FNIQSLYPLIKDPSLNLHIVLVRDPRAVFRSRETKGDLMTDSRIVWQHEQKLKE 237  
 Db 179 FNIQSLYPLIKDPSLNLHIVLVRDPRAVFRSRETKGDLMTDSRIVWQHEQKLKE 237  
 QY 242 PYVYMOVYCQSOLEIKYK--TIQSLPKALOERYLVREEDLARAPVQTSRYEFGLEFL 299  
 238 PYVYMOVYCQSOLEIKYK--TIQSLPKALOERYLVREEDLARAPVQTSRYEFGLEFL 296  
 Db 238 PYVYMOVYCQSOLEIKYK--TIQSLPKALOERYLVREEDLARAPVQTSRYEFGLEFL 296  
 QY 300 PHLQWVHNITRGKMGNDHAFHTNARDALNVSQAMRSLPEYKYSRLQKAGDANLLG 357  
 297 PHLQWVHNITRGKMGNDHAFHTNARDALNVSQAMRSLPEYKYSRLQKAGDANLLG 356  
 Db 297 PHLQWVHNITRGKMGNDHAFHTNARDALNVSQAMRSLPEYKYSRLQKAGDANLLG 356  
 QY 358 GYRVHVSDEQRNLDDLSTWVPEQI 386  
 357 GYRVHVSDEQRNLDDLSTWVPEQI 382  
 Db 357 GYRVHVSDEQRNLDDLSTWVPEQI 382

## RESULT 5

Q9GUP4 PRELIMINARY; PRT; 395 AA.

ID Q9GUP4  
 AC Q9GUP4  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE N-acetylglucosamine 6-O-sulfotransferase.  
 GN CHST5 OR I-GLOCAC-6-ST.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6; TISSUE=INTESTINE;  
 RA MBDLINE=99423499; PubMed=10491328;  
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;  
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-Sulfotransferase that is Highly Restricted to Intestinal Tissue."  
 RT Biochem. Biophys. Res. Commun. 263:543-549(1999).  
 DR EMBL: AF176841; AAD56002.1; -  
 DR EMBL: AF176840; AAD56002.1; -  
 DR MGD: MGI:1931825; Chst5.  
 KW Transferase.  
 SO SEQUENCE 395 AA; 44537 MW; 3PDE71E43ED383BE CRC64;

Query Match 49.9%; Score 1017.5; DB 11; Length 395;  
 Best Local Similarity 56.4%; Pred. No. 1.7e-85;  
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLLPK--KKKLLFLVSQMAILLALFFHMYSHNITSLSMKAOPERMHVIVLSSWRSGSSFFV 58  
 1 MRLPSSSTVMLSLVAGTGIIIVF--LVSRQVPS-SPAGIGRHHVHVIVLSSWRSGSSFFV 56  
 Db 1 MRLPSSSTVMLSLVAGTGIIIVF--LVSRQVPS-SPAGIGRHHVHVIVLSSWRSGSSFFV 56

QY	61	LEQGHDPVEYLMEBAMHWMFFKOSTWMLIMAVRDILRAFLCDMSVFPAYMEPCGRQ	130
Db	81	LFSGHPDVEFLMEBAMHWITLQSGSNAITLHMAYRDLMSIFLDDMDVPAYM-PQSRNL	139
QY	121	SSLFQWENSRLCSAPACDITPQDEILIPRACRLLCSQQPEVEYKRCRSTSHVLYKEVR	180
Db	140	SAFENMATSRCSPPACSAFPRGTISKQDCKLCTPQPSLAREACRSTSHVLYKEVR	199
QY	181	FENQSLTFLPKDPSLNLHIVLRDPRAFVRSRERTKGDIMDSRTVMQHEOKLEKD	240
Db	200	FENQVITFLPDLSDALNRIYHVRDRPRAVRSREAGPILARNGSYVLGTN-GKWEAD	258
QY	241	QPYVMQVYQCSQLEIYK--TIQSLPRALDORLYLVAYEDLARAIPVQTSRMYEEVLEF	258
Db	259	PHLRIRIRVCSHVRILDEAATLKP-PPLRLKRYLVLFVEDLARPELEIRALVAFETLTL	317
QY	299	LPHLOTWYHNITRCKGMGD--HAETHINARDALNVSQAMRMSLPYEKVSRLOKAGGDAMNL	356
Db	318	TPOLEAMHNHTTHSGSGKRPTEAETHSSRNARNVSOAMRHNLPTKILRVOEVACAGALQ	377
QY	357	LGYRHVRSEDEORNLDDL-----SIWYVPE 363	
Db	378	LGIRPVYSADQQRDLTDLVLRPRDPHRSNAPD 411	

ID Q9EP78 PRELIMINARY; PRT; 484 AA

DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE N-acetylglucosamine 6-0-sulfotransferase GST-5 (N-  
DE acetylglucosamine-6-0-sulfotransferase) (2600013M0R1K protein)  
DE (carbohydrate (N-acetylglucosamino) sulfotransferase 7).  
GN GST7 OR GST5 OR 2600013M0R1K.  
GN Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
CX NCBI\_TaxID=10090;

RA Bhakta S., Bartes A., Bowman K.G., Kao W.M.

R1 "Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2  
R2 (GST-5)." ;  
R3  
R4 J. Biol. Chem. 275:40226-40234 (2000).  
R5 [2]  
RN

RL Submitted (MAR-2000) to the EMBL/GenBank/DDI

RN 13  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
RX MEDLINE=21085660; Pubmed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arkawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,  
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamazaki I.,  
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito K.,  
RA Kaotaka K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Blake J., Boffelli D., Bojunga N., Carninci

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni U., Mashima J., Mazzarelli J., Mombaerts P.,  
Riedel G., Riedel T., Rodriguez C., Salomato N.



```

QY 101 VFLCDMSVFAVMEPCG-PRRQ-----SIFQWENSRLCSAPACDIIPQDE----I 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 160 LFRCDSEVLRLYAPPGDPAPARADTANLTALFRWRTNKVICSPPLCGAPARAEEVL 219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 IPRACHLLCSQOPFEVEKACRSYSHVYLKEVFNLSLYPLDKPSLNHTVLYVD 206
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 220 VEDTACRSQCPPAIRALECECKRYVYVYKDYRLDLGLVYPLRLDPGLNKLKVOGLFD 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 PRAVFRSREPTKGLMIDS-----RIVM-----GQHEOKLKKEDQPY 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 280 PRAVHNSRLKSGRGLLRESIQVLRTRQGRDFRFRVLAVGARGPQSGSALPAAPRADF 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 ----VMQVICOSELEIKTQSLPKALQERYLLVRYEDLARAPVQTSRYEFGLEFPD 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 340 FLTGALVEICEAWLRDLIFARGAPAWLRRLRYLRLEDYLRQPRQDLRLRLESGRLALA 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 HLOTWVHNITRKGKG-DHAFTNARDALNVSOQWRMSLPEYKVSRLQACGDAMNLGY 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 400 ALDAFALNMTRGAAVAGADREFHLASDAREAVAHMRERLSREQVRQVEACAPAMRLAY 459
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 RHVSEOE 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 460 --PRSGEE 465

RESULT 10
Q9NS84 PRELIMINARY; PRT: 486 AA.
AC Q9NS84;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN C6ST-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRNIN;
RA MEDLINE=20347214; PubMed=10781596;
RA Kitagawa H., Fujita W., Ito N., Sugahara K.;
RT "Molecular cloning and expression of a novel chondroitin 6-O-
RT sulfotransferase."
RL J. Biol. Chem. 275:21075-21080(2000).
DR EMBL: AB03187; BAE03217.1; -.
KW Transferase.
SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;

Query Match 29.7%; Score 605.5; DB 4; Length 486;
Best Local Similarity 37.0%; Pred. No. 2.1e-47;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

```

```

QY 301 HLOTWVHNITRKGKG-DHAFTNARDALNVSOQWRMSLPEYKVSRLQACGDAMNLGY 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 400 ALDAFALNMTRGAAVAGADREFHLASDAREAVAHMRERLSREQVRQVEACAPAMRLAY 459
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 RHVSEOE 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 460 --PRSGEE 465

RESULT 11
Q98276 PRELIMINARY; PRT: 530 AA.
AC Q98276;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE N-acetylglicosamine-6-O-sulfotransferase long form.
GN CHST2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WOLE EMBRYOS;
RC MEDLINE=98380482; PubMed=9712885;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan O., Kurosawa N.,
RA Mitsuoka C., Kanagaki R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglicosamine-6-O-
RT sulfotransferase."
RL J. Biol. Chem. 273:22577-22583(1998).
DR EMBL: AB011452; BAA32138.1; -.
DR EMBL: AB011452; BAA32139.1; -.
DR EMBL: AB011451; BAA32137.1; -.
DR MGD: MGI:1691160; Chst2.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer: 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113EB735C363BC CRC64;

Query Match 29.4%; Score 598.5; DB 11; Length 530;
Best Local Similarity 36.8%; Pred. No. 1.1e-46;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

```



```

DE (G1CNAC6ST).
GN GNS6T.
OC Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kanane T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsumura K., Kennadi R., Habuchi O., Yokoyama T., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.",
RL J. Biochem. 124:670-678(1998).
DR EMBL: AB014680; BAA34266.1;
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransferase.
KW Sulfotransferase.
SQ
SEQUENCE 531 AA; 56768 MW; D2EDB74E95B5162F CRC64;

Query Match      28.8%; Score 587.5; DB 4; Length 531;
Best Local Similarity 36.0%; Pred. No. 1.1e-45;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMVVLSSNRSGSSFGQLFGQHPDVFYLMEDPAHVMWTFKOSTAMMLHAAVRLIRA 100
DB 164 KRHMVYFTTWSSGSSFFGELFNQNEVFLPEPVWVWVKLYPGDAVSLQGAARDMLSA 223
QY 101 VFLDMSVFAVMEPPGRRQS---SLFQWNSRALCSAPACDIIPODEIIPAHORLL 155
DB 224 LYKDLSEVFOGLYPAGSGGNLTLLGIFGATKKNVCSSPLCPAY-RKEVYGLVDRVCK 282
QY 156 -CSQPEVEVEKACRSYSHVYLKEVRFNLQSLYPLKDPDLMLHIVHLVDRPRAVFRSR 214
DB 283 KCPQRLAREFEEDCRKRYTLVIGKGVFVAVLAPLLRDLALDLKVIHLVDRPRAVASSR 342
QY 215 ERTKGLMIDSRIMY-----GQHEQKIKKE---DQPYV---VMQVICS 252
DB 343 IRSHGILRESLOVVRSDPRANRMPLEAGHKLGAKEGAGVGADPADYHLAGMEVICS 402
QY 253 QLEIYKTIQSLPRALQERYLLVRYEDLARAPVQTSRMVEFGLEFLPHLOTVWNIITRG 312
DB 403 MAKLTQIALQPPDLQGHYLVRYEDLVGPVXTLRVYDFVGLVSPEREKALWMTSG 462
QY 313 KGMGDHAFHTNARDALNVSGAQRMSLPYEKYSRLQKAGDAMNLGTRHYRSFQDQRL 372
DB 463 SSSSKPFVVSARNAIQOANAMRTALTFOQIKQVEFCYQPMVAVLGERVNSPEEVKDL 522
QY 373 LDLL 376
DB 523 KTL 526

RESULT 15
Q9E0C0 PRELIMINARY; PRT; 411 AA.
AC Q9E0C0;
DT 01-MAR-2001 (TREMBLE). 16, Created)
DT 01-MAR-2001 (TREMBLE). 16, Last sequence update)
DT 01-DEC-2001 (TREMBLE). 19, Last annotation update)
DE Galactose 6-O-sulfotransferase GST-1.
GN CHST1 OR GST1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10050;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=MAMMARY GLAND;
RA Hemmerlich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,

```

```

RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RT family.";
RT Glycobiology 11:75-87(2000).
RL EMBL: AF280087; AAG48245.1;
DR MGD; MGI:1924219; Chst1.
DR Transferrase.
KW Transferrase.
SQ
SEQUENCE 411 AA; 46903 MW; B1AE590EF5B9CBDC CRC64;

Query Match      27.4%; Score 559; DB 11; Length 411;
Best Local Similarity 35.9%; Pred. No. 3.4e-43;
Matches 132; Conservative 65; Mismatches 135; Indels 36; Gaps 10;

QY 27 YSHNLSLSMKAPPRMHLVLSNRSGSSFGQLFGQHPDVFYLMEDPAHVMWTF----F 82
DB 53 FSYNLS-----RRTVLIATTRSSGSSFFGQLFNQHMVFYLFLEPLVHVNLTLLPRF 104
QY 83 KQSTA-----WMLHMAVRLDIRAVFLCDMSVPDAVMEPPGRRQS--LFWQNSRALCSAP 136
DB 105 TQKSPADRNVHGLASRDILNSLYDCDLYFLENTIKPPVNHITNRVRRGASRYLCSR 164
QY 137 ACDIIP---QDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRFNLQSLYPLKD 193
DB 165 VCD--PPGSSDLILEGDCVRMCGLLNLTLLAELACRERSHVAIKTVRVENVNDLRALYED 222
QY 194 PSLNHLIVHLVDRPRAVFRSRERTKGDLMIDSRIMYQHEQKIKKEDQPYV---KQVI 249
DB 223 PRLNKKVIQLVDRDPGILASRSETFRDYRLMRVMYGTGR-----KPNLDVYQTLTV 275
QY 250 COSQLEIYKTIQSLPRALQERYLLVRYEDLARAPVQTSRMVEFGLEFLPHLOTVWNI 309
DB 276 CEDFSSSVSTGLMRPSWLGKMYLVRYEDLARNPYKKEELIYFELGIDLDHVAHWIINN 335
QY 310 TFG-KMGDHAFTNARDALNVSGAQRMSLPYEKYSRLQKAGDAMNLGTRHYRSFQDQ 368
DB 336 TFGDPTLGHKKYST-VNSAATAEKWRPRLSYDIAFAQNAQOQVLAQDGYKMANSEBEL 394
QY 369 RNLDDL 376
DB 395 KNPALSLV 402

```

Search completed: January 11, 2003, 01:20:58  
Job time : 93 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:19:36 (Search time 24 Seconds  
(without alignments)  
473.219 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038

Sequence: 1 MLPRKMKLLFLVSQMAIL.....EQRLNLLDLSTWVPEQIH 386

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5a\_COMB.pep:\*

2: /cgn2\_6/prodata/1/1aa/5b\_COMB.pep:\*

3: /cgn2\_6/prodata/1/1aa/5c\_COMB.pep:\*

4: /cgn2\_6/prodata/1/1aa/5d\_COMB.pep:\*

5: /cgn2\_6/prodata/1/1aa/5e\_COMB.pep:\*

6: /cgn2\_6/prodata/1/1aa/5f\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	4	US-09-045-284A-2
2	2038	100.0	386	4	US-09-190-911-1
3	598.5	29.4	483	3	US-09-263-023-2
4	598.5	29.4	483	3	US-09-471-867-2
5	587.5	28.8	484	3	US-09-263-023-4
6	587.5	28.8	484	4	US-09-471-867-4
7	549	26.9	411	4	US-09-013-188-2
8	500.5	24.6	458	2	US-08-555-878-2
9	482	23.7	479	2	US-08-999-514-2
10	94.5	4.6	848	4	US-09-540-824-27
11	92	4.5	359	3	US-09-150-133-11
12	92	4.5	359	3	US-09-150-141-11
13	92	4.5	359	4	US-09-374-493-11
14	92	4.5	359	4	US-09-374-824-11
15	92	4.5	359	4	US-09-374-824-11
16	87.5	4.3	380	3	US-09-150-133-9
17	87.5	4.3	380	3	US-09-150-141-9
18	87.5	4.3	380	4	US-09-374-493-9
19	87.5	4.3	380	4	US-09-374-824-9
20	87.5	4.3	380	4	US-09-374-824-9
21	86.5	4.2	831	1	US-08-073-384C-5
22	86.5	4.2	831	1	US-08-254-359A-5
23	86.5	4.2	831	1	US-08-483-043-5
24	86.5	4.2	831	1	US-08-481-238-5
25	86.5	4.2	831	2	US-08-471-066B-5
26	86.5	4.2	831	2	US-08-484-956-5
27	86.5	4.2	831	2	US-08-757-653-5

28	86.5	4.2	831	2	US-08-599-491-5	Sequence 5, Appl1
29	86.5	4.2	831	2	US-08-756-386-5	Sequence 5, Appl1
30	86.5	4.2	831	2	US-08-823-516-5	Sequence 5, Appl1
31	86.5	4.2	831	3	US-08-682-853A-5	Sequence 5, Appl1
32	86.5	4.2	831	3	US-08-759-036-5	Sequence 5, Appl1
33	86.5	4.2	831	3	US-08-758-314-5	Sequence 5, Appl1
34	86.5	4.2	831	3	US-09-350-309-5	Sequence 5, Appl1
35	86.5	4.2	831	4	US-08-520-946-5	Sequence 5, Appl1
36	83	4.1	222	2	US-08-485-721-9	Sequence 9, Appl1
37	83	4.1	222	2	PCT-US93-08326-9	Sequence 9, Appl1
38	83	4.1	222	2	US-08-392-938-9	Sequence 9, Appl1
39	83	4.1	920	3	US-08-930-996A-8	Sequence 8, Appl1
40	82.5	4.0	1140	3	US-08-471-112A-4	Sequence 4, Appl1
41	82.5	4.0	1809	3	US-09-012-515A-12	Sequence 12, Appl1
42	82.5	4.0	1809	4	US-08-360-144A-12	Sequence 12, Appl1
43	82.5	4.0	1809	4	US-09-012-504A-12	Sequence 3, Appl1
44	82.5	4.0	2549	4	US-08-471-112A-3	Sequence 12, Appl1
45	82.5	4.0	2549	5	PCT-US93-06722-12	Sequence 12, Appl1

ALIGNMENTS

RESULT 1

US-09-045-284A-2

Sequence 2, Application US/09045284A

Patent No. 6265192

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107051

CURRENT APPLICATION NUMBER: US/09/045,284A

CURRENT FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 386

TYPE: PRT

ORGANISM: Homo sapiens

US-09-045-284A-2

Query Match	100.0%	Score 2038	DB 4	Length 386
Best Local Similarity	100.0%	Pred. No. 2	2e-212	
Matches 386	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MLPRKMKLLFLVSQMAILALFFHMYSHNTSLSKRAQPERKHVLTLSMRSGSFFVGO	60	
DB	1	MLPRKMKLLFLVSQMAILALFFHMYSHNTSLSKRAQPERKHVLTLSMRSGSFFVGO	60	
QY	61	LFQGHDPVLYLMEPAHVWMTFKQSTAMMLHNAVRLDILAVFLCDMSVFDAYMEGPFRQ	120	
DB	61	LFQGHDPVLYLMEPAHVWMTFKQSTAMMLHNAVRLDILAVFLCDMSVFDAYMEGPFRQ	120	
QY	121	SSLFQWENSALCSAPACDIIPDDETIPRAHCRILCSQOPEFVEVERACGSHVYKEVR	180	
DB	121	SSLFQWENSALCSAPACDIIPDDETIPRAHCRILCSQOPEFVEVERACGSHVYKEVR	180	
QY	181	FFNLQSLYPLKDPKSLNLHIVLVRDPRAVFRSRETKGDLMDISRIYVGOHQKLRKED	240	
DB	181	FFNLQSLYPLKDPKSLNLHIVLVRDPRAVFRSRETKGDLMDISRIYVGOHQKLRKED	240	
QY	241	QPYVWQVICOQSOLEYTKTQISLPRALQSRVLLVREDILARAVQTSMTFFVGLFFP	300	
DB	241	QPYVWQVICOQSOLEYTKTQISLPRALQSRVLLVREDILARAVQTSMTFFVGLFFP	300	
QY	301	HLCITWENITRGKMGDAFHTNARDALNVSQAWKWSLPEYKVSRLQKACGDMNLGGR	360	
DB	301	HLCITWENITRGKMGDAFHTNARDALNVSQAWKWSLPEYKVSRLQKACGDMNLGGR	360	
QY	361	HVSRQEQRLNLLDLSTWVPEQIH	386	
DB	361	HVSRQEQRLNLLDLSTWVPEQIH	386	

Db 361 HVHSEQEQRLNLLDLSTWTPVPEQIH 386

RESULT 2

US-09-190-911-1

; Sequence 1, Application US/09190911

; Patent No. 6365365

; GENERAL INFORMATION:

; APPLICANT: Bistrup, Annette

; APPLICANT: Rosen, Steven D.

; APPLICANT: Tangemann, Kirsten

; APPLICANT: Hemmerich, Stefan

; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

; FILE REFERENCE: 6510-107CIP

; CURRENT APPLICATION NUMBER: US/09/190,911

; CURRENT FILING DATE: 1998-11-12

; EARLIER APPLICATION NUMBER: 09/045,284

; EARLIER FILING DATE: 1998-03-20

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 386

; TYPE: PR1

; ORGANISM: H. sapiens

US-09-190-911-1

Query Match 100.0%; Score 2038; DB 4; Length 386;  
Best Local Similarity 100.0%; Pred. No. 2,2e-212;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPRKKKLLFLVSCNAILALFFHNTSHNLSLSKKAQPERNHVLYSSMRSGSSFFVQ 60

Db 1 MLPRKKKLLFLVSCNAILALFFHNTSHNLSLSKKAQPERNHVLYSSMRSGSSFFVQ 60

QY 61 LFGHPVFTLMEPAWMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYVEPPRRQ 120

Db 61 LFGHPVFTLMEPAWMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYVEPPRRQ 120

QY 121 SLPFOWNSRALCSAPACDIIPODEIIPRAHCRLLCSQCFEVEYKACSYSHVYLKEVR 180

Db 121 SLPFOWNSRALCSAPACDIIPODEIIPRAHCRLLCSQCFEVEYKACSYSHVYLKEVR 180

QY 181 FPNLOS:YPLLKPSLNLH:VHLVDRPRAVFRSRETKGGLMDSKRYVMQGHOKLKED 240

Db 181 FPNLOS:YPLLKPSLNLH:VHLVDRPRAVFRSRETKGGLMDSKRYVMQGHOKLKED 240

QY 241 QPYVVMQVICOQSOLEIKTQSLPKALOERYLLVRYEDLARAVAOQTSRMVEFVGLFEP 300

Db 241 QPYVVMQVICOQSOLEIKTQSLPKALOERYLLVRYEDLARAVAOQTSRMVEFVGLFEP 300

QY 301 HLOTWVNITRGKMGDHAFTINARDAALNVSQAMRWSLPYEKYSRLQKACGDAMNLLGR 360

Db 301 HLOTWVNITRGKMGDHAFTINARDAALNVSQAMRWSLPYEKYSRLQKACGDAMNLLGR 360

QY 361 HVHSEQEQRLNLLDLSTWTPVPEQIH 386

Db 361 HVHSEQEQRLNLLDLSTWTPVPEQIH 386

RESULT 3

US-09-263-023-2

; Sequence 2, Application US/09263023

; Patent No. 6037159

; GENERAL INFORMATION:

; APPLICANT: Uchimura, Kenji

; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadomatsu, Kenji

; APPLICANT: Kannagi, Reiji

; APPLICANT: Habuchi, Osami

; APPLICANT: Muramatsu, Takashi

; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

; FILE REFERENCE: TOYAMA1.001AUS

; CURRENT APPLICATION NUMBER: US/09/263,023

; CURRENT FILING DATE: 1999-03-05

; EARLIER APPLICATION NUMBER: JP 10-54007

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: JP 10-177844

; EARLIER FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 483

; TYPE: PR1

; ORGANISM: Mus musculus

US-09-263-023-2

Query Match 29.4%; Score 598.5; DB 3; Length 483;  
Best Local Similarity 36.8%; Pred. No. 3e-56;  
Matches 134; Conservative 64; Mismatches 177; Indels 29; Gaps 6;

QY 41 ERMHVLVSSMRSGSSFFVQGLFGQHPDVFTLMEPAWMTFKQSTAMMLHMAVRDLIRA 100

Db 116 KROLVVEFTWRSQSSFFGELFNQNPVEVFLYEPVHWOKLYPDGDAVSLQGAARDMLSA 175

QY 101 VFLCDMSVFDAYMEPPRRQS----SLFOWNSRALCSAPACDIIPODEIIPRAHCRLL- 155

Db 176 LYRCDISVFOLYSPAGSGGRNITTLIGTGAATNKYVCSPLCPAY-KREYVGLVDDRVC 234

QY 156 -CSQCFEVEYKACSYSHVYLKEVRFNLSQSLYPLKPSLNLH:VHLVDRPRAVRSR 214

Db 235 KCPORLARFEDECRKRYTVIKGVRFDAVAVLAPLLKDPALDLKVLHVLDRPRAVASSR 294

QY 215 EFRKDDLMDSRYV-----QQHEQKLEKEDQ----PYV--VWQVCOQS 252

Db 295 ITRSGRLRESQVYRSDDPRAHMRPLEAAGHKLGAKBEQMGGRADYHALCAEVLGNS 354

QY 253 QLEIKTQSLPKALOERYLLVRYEDLARAVAOQTSRMVEFVGLFEPHLOTWVNITRG 312

Db 355 NAKTQTLQRPDMLQGHVLYVRYEDLVGDPYKTLRYVYDEVGLVSPMEQFALNMTSG 414

QY 313 KMGDHAFTINARDAALNVSQAMRWSLPYEKYSRLQKACGDAMNLLGRHVHSEQEQRL 372

Db 415 SSSSKRPVVASRNATQANAMRALTFOIKQVEFCYOPMAVGLYERVNSPEEVKOLS 474

QY 373 LDLL 376

Db 475 KTL 478

RESULT 4

US-09-471-867-2

; Sequence 2, Application US/09471867

; Patent No. 6455289

; GENERAL INFORMATION:

; APPLICANT: Uchimura, Kenji

; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadomatsu, Kenji

; APPLICANT: Kannagi, Reiji

; APPLICANT: Habuchi, Osami

; APPLICANT: Muramatsu, Takashi

; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

; FILE REFERENCE: TOYAMA1.001AUS

; CURRENT APPLICATION NUMBER: US/09/471,867

; CURRENT FILING DATE: 1999-12-23

; EARLIER APPLICATION NUMBER: US 09/263,023

; EARLIER FILING DATE: 1999-03-05

; EARLIER APPLICATION NUMBER: JP 10-54007

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: JP 10-177844

; EARLIER FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 483



```

; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

```

Query Match	29.4%;	Score 598.5;	DB 4;	Length 483;
Best Local Similarity	36.8%;	Pred. NO. 3e-56;		
Matches 134; Conservative	64;	Mismatches 137;	Indels 29;	Gaps 6

[illegible]

```

RESULT 5
US-09-263-023-4
? Sequence 4, Application US/09263023
? Patent No. 6037159
? GENERAL INFORMATION:
? APPLICANT: Uchimura, Kenji
? APPLICANT: Muramatsu, Hideki
? APPLICANT: Katomatsu, Kenji
? APPLICANT: Kanaghi, Reiji
? APPLICANT: Habuchi, Osami
? APPLICANT: Muramatsu, Takashi
? TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
? TITLE OF INVENTION: DNA ENCODING THE SAME
? TITLE OF INVENTION: TOYAMA41.001AUS
? FILE REFERENCE: TOYAMA41.001AUS
? CURRENT APPLICATION NUMBER: US/09/263.023
? CURRENT FILING DATE: 1999-03-05
? EARLIER APPLICATION NUMBER: JP 10-54007
? EARLIER FILING DATE: 1998-03-05
? EARLIER APPLICATION NUMBER: JP 10-177844
? EARLIER FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 10
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 4
? LENGTH: 484
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-09-263-023-4

```

Query Match:	28.8%	Score 587.5	DB: 3	Length 484;
Best Local Similarity:	36.0%	Pred: 14.6e-55;		
Matches 131;	Conservative 67;	Mismatches 177;	Indels 29;	Gaps 6

  

QY	41	ERMHVLVLSWRSGSFFVQLFGQEDHVFYLMERPMHWMTFKOSTMTMLHMAVRLIRA	100
b	117	KRHMYVFTTWRSQSGFEELNQNDEVFILPEPMHWQKLRGCDVSLVSGARQDLWSLA	176

```

QY 101 VELDMASVCEADYVMPGPRBROS-----SLFQWENSNALCSAPACDILPDELIIPAHORCLL- 155
Db 177 LYRCDLSVFLYTSVAGSGGRRLTLGLFGFAATNKVYVSSPLCEAY-RKEVYGLVDYDRVC 225
QY 156 -CSOOPEEVEVAKCRASYHVVLKCVREFEYNQSLYPLIKDESLNHLVHLVDRPRAVFSR 214
Db 236 KCPQORLARPEEBCRKRYTLIKGVBEVDAVLAFLPRLDRBALDKVHLVDRPRAVASSR 295
QY 215 ERTKGDMLDIBERIYV-----GQHQKRLKE-----DQPYV---AMOVICOS 252
Db 296 IRRSHGLIRRELQYVRSRDPRAHMRPLEAGHGLAKKEGVSGPRAVYHNLGAMVEYICNS 355
QY 253 QLEIYKTIQSLPKALOERYLLVRYEDLARAPVAOTSMMYEFGLEFLPHIQTVHNHTTG 312
Db 356 MAKPLQIALALQPPDMLQGHYLVREDYDVJQPVKTLRRYVYFGVGLVSPREDEQALNNTSG 415
QY 313 KGMDEHAFHTAPDALVYNSQAMRSLYEKVSRLQKACGDAANLILGVRHRSQEOQNTL 372
Db 416 SGSSSKRPVVSARNAQTGAANMRFALLFQOLKYQEEDCYCPMAVLQYERANSPEVEYKDL 475
QY 373 LDLL 376
Db 476 KTL 479

```

```

: RESULT 6
: US-09-471-867-4
: Sequence 4, Application US/09471867
: Patent No. 645289
: GENERAL INFORMATION:
: APPLICANT: Uchimura, Kenji
: APPLICANT: Muramatsu, Hideki
: APPLICANT: Kadomatsu, Kenji
: APPLICANT: Kanmaji, Reiji
: APPLICANT: Hachuchi, Osami
: APPLICANT: Muramatsu, Takashi
: TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
: FILE OF INVENTION: DNA ENCODING THE SAME
: FILE REFERENCE: TOYAMA1 001AUS
: CURRENT APPLICATION NUMBER: US/09/471,867
: CURRENT FILING DATE: 1999-12-23
: EARLIER APPLICATION NUMBER: US 09/263,023
: EARLIER FILING DATE: 1999-03-05
: EARLIER APPLICATION NUMBER: JP 10-54007
: EARLIER FILING DATE: 1998-03-05
: EARLIER APPLICATION NUMBER: JP 10-177844
: EARLIER FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 484
: TYPE: CDS
: ORGANISM: Homo sapiens
: US-09-471-867-4

```

	Query Marc36	28.0%	Score 587.5	DB 4	Length 484
	Best Local Similarity	36.0%	Pred. No. 4.6e-55		
	Matches 131	Conservative 67	Mismatches 137	Indels 29	Gaps 6
QY	41 ERMHVLSSMRSGSSFFVQQLFGQHFDFYLLMEPAWHYMTFKOSTAMLMHAYRDLIRA	100			
Db	117 KRHMVYETITWRSGSSFFGELEQNQNEVEFELYEPWVHMQKLYPGDAVSLQGAARDLISA	176			
QY	101 VFLCDMSVEPAHYAEPPGRKRS-----SLFQWENSRKLCAPACDIIIPQDEIIFRACGLL-	155			
Db	177 LYRDLISVFPQLYSPASSGGGRNLTLLGIFCAATNKKVCCSSPLCPAY-RKEVYGLVDKRVCK	235			
QY	156 -CSQGPFEVEYERKASRSHVYLKEVFNFLLQSLYLLDPSLNIIHYHVLDDPAVYRSR	214			
Db	236 KCPQRLAREEDECRCRITLIVIKGVAFVDVAVLADLPDADLKLIVLVDVDPAAVASSR	235			
QY	215 ERTGGLMDISRIVM-----GGHEOKLKE-----DQPY--VWQVICOSS	252			

```
Db 296 IRSRHGLRESLOVSRDPRAHMPLEBAHGKLGAKKEGVGPADYHALGAMEVICS 355
Qy 253 QLEIVKIQSLPALOEIRYLIVYEDLARPAVQTSRMKEFEVLEFLPHLQTVHNITG 312
Db 356 MXTLQALPDPMLQGHVLYRELDVGPVTLRRVYDEVGLLPSFMEQPALMTG 415
Qy 313 KGMGDHAFHTNARDALVNSQAMRWSLPEKYSRLQKACGDAMNLGVRHVSDEQRNLL 372
Db 416 SGSSSKPEFVVSANRATGAAMAMTALTFOQIKVEEFCQPMNAVLYERNVNSPEVKDLS 475
Qy 373 LDLL 376
Db 476 KTL 479

RESULT 7
US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 639938
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TIME OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 26.9%; Score 549; DB 4; Length 411;
Best Local Similarity 34.3%; Pred No. 5, 4e-51;
Matches 133; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

Qy 6 KMKLLFLVS--OMALLAFHMYSHNISLSMKAQPERM-----HVLV 47
Db 6 KAVLLALASIAIOYTAIRFTAKSFHTCPGLAAGLAELCESEPTFAYNLSRKTHTLI 65
Qy 48 LSSWRSSSVFQGLFGQHPDVFYLMPEAMHVMWT----FKOSTA----MMLNAVDDL 99
Db 66 LATTRSSSVFQGLFQNDLVFYLFEPLVHVQNTLIPFTQKSPADRRVMGASRDLLR 125
Qy 100 AVFLCDMSVDAWMEGPPRRSS--LEQWENSRLCSAPACDIT--PQDITIRAHCRILC 156
Db 126 SLVDCDLYFLENYIKPPVNHHTDRIFRGASVILCSRPVCDPPADIVLEGGDCVRKC 185
Qy 157 SQCPFEVEKACRSYSHVYLKEVRFNLSQSLYPLKDPSLNHTIYLVDPRAVFNRSR 216
Db 186 GLNLTVAAACRERSHVALKTIVPEVNDLRALVEDPRLNLKVIQLVDPDRIILASRSE 245
Qy 217 TKGLMDSIRIVGQHEOKLKEKEDOPYV---MOVICQSLEIYKTIQSLPKALOERYL 272
Db 246 TFRDTYLMWLMWGTG-----KRYNDVYQTLTVCEDEFSNSYSTGLMRPWLKGYM 298
Qy 273 LVRVEDARPAVQTSRMKEFEVLEFLPHLQTVHNITG--KMGDHAFTNARDALVNS 331
Db 299 LVRVEDLARPMKRTTEIIGFQICPLDSHVARWIQNNTRGDPTLGHKKYGT--VRNSATA 357
Qy 332 QAMRWSLPEKYSRLQKACGDAMNLGVRHVSDEQRNLLDL 376
Db 358 EKMFRSLYDIAFAQVACQVLAQLGYKIASEBELKNPSVSLV 402

RESULT 8
US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
```

```
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,878
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-655-878-2

Query Match 24.6%; Score 500.5; DB 2; Length 458;
Best Local Similarity 33.2%; Pred. No. 1, 1e-45;
Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

Qy 34 LSMKAQPERMIVLYLSSWRSSSVFQGLFGQHPDVFYLMPEAMHVM--WMTEKOSTAMWLH 91
Db 104 LGIAPEPRRHVLLMATIRTGSSFVGEFNFQGNIFYLFEPLMIKERVYTFEPGANNVG 163
Qy 92 MAV--RDLIRAVFLCDMSVDAWMEGPPRRQ--SSLQWENSRLCSAPACDITPQ--DEI 146
Db 164 SALVYRDVQLQILLCDLYLLESFISPADEHLLTALFRGSSHSICEPPVC--TPSLKY 221
Qy 147 IPRAHCR--LLCSQCPFEVEKACRSYSHVYLKEVRFNLSQSLYPLKDPSLNHTIYLV 205
Db 222 FEKYHCKNRRCGPLNITLAEACRKRQHMALKTIVRIQLEFLQPLAEDPRLDLTIQLVR 281
Qy 206 DPRAVFRSRETKGDLMDISRIWQHEOKL-----KEDQPYVVMQVICS--QLE 255
Db 282 DPRALVSR-----NV---AFSGKYESWKMAEGEAPLQDEE--VQRLNGCESIRLS 330
Qy 256 IYKTIQSLPKALOERYLVYRELDARAPVQTSRMKEFEVLEFLPHLQTVHNITRKG 315
Db 331 AELGRQ--PRMLRGHYMLVREYVARAPLRKALEYRPAGIHPTPYQVEWVRANTQAP--Q 388
Qy 316 GDHAFHTNARDALVNSQAMRWSLPEKYSRLQKACGDAMNLGVRHVSDEQRNLLDL 375
Db 389 DSNGLIYTKNSSEQFEKMRRESIPFKLAQVYODACERPMRLFGYKILASSAQELINRSL 448
Qy 376 L 376
Db 449 L 449

RESULT 9
US-08-899-514-2
```

```

Sequence 2 Application US/08899514
Patent No. 5910581
GENERAL INFORMATION:
APPLICANT: HABUCHI, OSAMI
APPLICANT: FUKUTA, MASAKAZU
TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
TITLE OF INVENTION: FOR THE POLYPEPTIDE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-514-2

```



Mon Jan 13 09:14:05 2003

us-09-816-825-2.raii

Page 7

```

Query Match 4.5%; Score 22; DB 4; Length 359;
Best Local Similarity 19.5%; Pred. No. 0.14; Indels 98; Gaps 15;
Matches 63; Conservative 48; Mismatches 114;

QY 52 RSGSSFGVQLGCGHNDV---FYLMEPAMHWMFPGKSTAMMLEHVAVDLIRAVFLCDMS 107
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 95 RSGTILMALIIDAHPDVCGCGFTMLPSFLFWQAGWRND-WVNSGCI-----TQE 143
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 108 VFDAIMEGPRQSSLFWEMNSRALCSAPACDITQ-DEIIRAHRCILSCQPREVEYK 166
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 144 VFDDVV-----SAPFTEIYAKHSELAPR-----LCNKDP----- 172
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 167 ACSRSYHVLKEVREFNQSLPYPLKIDPSLNIHIVHLRDPRAVERREPRK---GDLM 222
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 173 ---YIALMLPIR-----RLRP-----NAKFLIMRDARAAYVHSIEKVPAGINT 216
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 223 IDSRIVMGQHEQKCLKKEDQPYVMQVLCQSQ---LEIYKTITQSLPKALOERYLLVRED 278
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 217 SDEISMFGVQMDLELR-----MTFCQNNAPAGOCIKY-----YER 251
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 279 LARAVAAOTSMRYEFGLEFLPHLOTWVHNITRGKMGCDHAFH-TNARDALVSGAMRMS 337
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 252 LIQKAEELIRITNLEFLDPESQOMLRHODLLGDEVYLDNDERSASQVKNSTNTALISWF 311
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 338 LPY-EKVSRLQKACGDAMNLGY 359
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 312 DCFSEETLRKLLDVAAPFLGILGY 334
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 15
US-09-374-492-11
: Sequence 11, Application US/09374492
: Patent No. 6207432
: GENERAL INFORMATION:
: APPLICANT: The Board of Regents of the University of Oklahoma
: TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
: FILE REFERENCE: 5820.545
: CURRENT APPLICATION NUMBER: US/09/374,492
: CURRENT FILING DATE: 1999-08-13
: EARLIER APPLICATION NUMBER: 09/150,141
: EARLIER FILING DATE: 1998-09-09
: EARLIER APPLICATION NUMBER: 60/072,994
: EARLIER FILING DATE: 1998-01-29
: EARLIER APPLICATION NUMBER: PCT/US99/16750
: EARLIER FILING DATE: 1999-07-23
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: wordperfect 8.0 (saved in ASCII format)
: SEQ ID NO 11
: LENGTH: 359
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
: US-09-374-492-11

```

Query Match	4.5%	Score 92	DB 4	Length 359
Best Local Similarity	19.5%	Prod. No. 0.14		
Matches	63	Conservative	48	Mismatches 114
				Indels 98
				Gaps 15
QY	52	RGSSGVSQLEQANDV---	FTYMEPAWHVMYTFKOSTAMWLHNAVRLDILRAVFLCDMS	107
		:::	:::	
		:::	:::	
Db	95	RSQTLLRAILDADHDVRCGGFTMLPSLWQAGRND--WVNNSGI-----	-TQE	143
QY	108	VFDALMEGRPRQSSLFQWENSRALCSAPACITIQ--DEIIRACRLLCQQPFVEVER		166
		:::	:::	
		:::	:::	
Db	144	VFDQAV-----	SATETIVAKHSELAR-----LCNKDP-----	172
QY	167	ACRSYHVLYKEVRFNFQSLTYPLKDSLSNLIHVALVRAVRSRPRK---GDLN		222
		:::	:::	
		:::	:::	
Db	173	---YTLMLPIR-----RLYP-----	NKFLIMTIDARAAYVHSMLEKRVPAVGYNT	216
QY	223	IDSLRVGCHCEOKLKKEDQPIYVMQVICSQ---DEIYKTQSPLKALQERLLVRYED		278
		:::	:::	
		:::	:::	
Db	217	SDEISMFWQWMDLKK---MTFOCCNNAAGOCITKYV-----YER		251

QY 279 LAKAPVQTSMEYEGLEFLPHLQTWHTNTRKCGMDAHF-TNRKDLNLSQAMRS 337  
Db 252 LIQPAEETLITNFDLPESQMRHODLIGDEVLDLDOEFSSQYKNSINTKALTSF 311  
QY 338 LPY-EKVSRLQACGADAMNLGY 359  
Db 312 DQFEETLRKLDVAFLGILGY 334

Search completed: January 11, 2003, 01:22:24  
Job time : 28 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:21:07 ; Search time 18 seconds  
(without alignments)  
416.045 Million cell updates/sec

Title: US-09-816-825-2  
Perfect score: 2038  
Sequence: 1 MLPPKKKKLLFLVSGMAI.....EQRLNLLDLSTWTEQI 386

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/pubppa/US08\_NEM\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEM\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEM\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEM\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_NEM\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_NEM\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_NEM\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	9 US-10-007-262-1	Sequence 1, Appl
2	2038	100.0	386	10 US-09-816-825-2	Sequence 2, Appl
3	1028	50.4	418	10 US-09-927-602-5	Sequence 5, Appl
4	1019	50.0	395	10 US-09-927-602-2	Sequence 2, Appl
5	1013.5	49.7	395	10 US-09-927-602-3	Sequence 3, Appl
6	1008	49.5	390	10 US-09-827-602-4	Sequence 4, Appl
7	821	40.3	171	10 US-09-927-602-8	Sequence 8, Appl
8	587.5	28.8	531	10 US-09-833-790-255	Sequence 255, App
9	519.5	25.5	169	10 US-09-927-602-6	Sequence 6, Appl
10	511.5	25.1	169	10 US-09-927-602-7	Sequence 7, Appl
11	336	16.5	179	10 US-09-927-602-9	Sequence 9, Appl
12	335.5	16.5	179	10 US-09-927-602-10	Sequence 10, Appl
13	293	14.4	174	10 US-09-827-602-11	Sequence 11, Appl
14	83	4.4	320	10 US-09-834-122-20	Sequence 20, Appl
15	88	4.3	2209	9 US-09-902-941-1903	Sequence 1903, Ap
16	88	4.2	2209	9 US-09-849-626-1903	Sequence 1903, Ap
17	86.5	4.2	831	9 US-10-033-297-5	Sequence 5, Appl
18	86.5	4.2	831	9 US-10-081-806-5	Sequence 5, Appl
19	86	4.2	824	10 US-09-801-368-312	Sequence 312, App

20	81	4.0	295	9 US-09-965-529-31	Sequence 31, Appl
21	81	4.0	1100	10 US-09-815-242-11861	Sequence 11861, A
22	79.5	3.9	177	8 US-08-424-5508-53	Sequence 53, Appl
23	79	3.9	402	10 US-09-815-242-11350	Sequence 11350, A
24	78.5	3.9	402	10 US-09-925-300-1612	Sequence 1612, Ap
25	78.5	3.9	745	9 US-09-844-988-10	Sequence 10, Appl
26	78.5	3.9	745	10 US-09-796-872-2	Sequence 2, Appl
27	78.5	3.9	745	10 US-09-844-908-10	Sequence 5356, Ap
28	78.5	3.9	899	10 US-09-815-242-13556	Sequence 12615, A
29	78.5	3.9	932	10 US-09-815-242-12615	Sequence 9, Appl
30	78	3.8	244	10 US-09-911-826A-19	Sequence 5460, Ap
31	78	3.8	1160	10 US-09-815-242-5480	Sequence 12451, A
32	77.5	3.8	1168	10 US-09-815-242-12151	Sequence 5373, Ap
33	77.5	3.8	1127	10 US-09-815-242-5373	Sequence 10240, A
34	77.5	3.8	1158	10 US-09-815-242-12522	Sequence 466, App
35	77	3.8	575	10 US-09-815-242-10240	Sequence 40, Appl
36	76	3.7	500	10 US-09-731-872-466	Sequence 13608, A
37	75.5	3.7	748	10 US-09-801-574-40	Sequence 19, Appl
38	75.5	3.7	1285	10 US-09-982-091A-2	Sequence 118, App
39	75	3.7	1179	10 US-09-815-242-13608	Sequence 304, App
40	74.5	3.7	666	9 US-09-331-631A-3	Sequence 144, App
41	74.5	3.7	755	10 US-09-881-852-19	
42	74	3.6	538	9 US-09-921-496-118	
43	74	3.6	538	10 US-09-874-923-118	
44	74	3.6	559	9 US-09-895-913A-304	
45	74	3.6	628	10 US-09-881-752A-144	

## ALIGNMENTS

RESULT 1					
US-10-007-262-1					
; Sequence 1, Application US/10007262					
; Patent No. US20020164748A1					
GENERAL INFORMATION:					
; APPLICANT: Bistrup, Annette					
; APPLICANT: Rosen, Steven D.					
; APPLICANT: Tangemann, Kirsten					
; APPLICANT: Hemmerlich, Stefan					
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3					
; FILE REFERENCE: 6510-107CIP					
; CURRENT APPLICATION NUMBER: US/10/007,262					
; CURRENT FILING DATE: 2001-11-08					
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911					
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12					
; NUMBER OF SEQ ID NOS: 8					
; SOFTWARE: FastSeq for Windows Version 3.0					
; SEQ ID NO 1					
; LENGTH: 386					
; TYPE: PRT					
; ORGANISM: H. sapiens					
US-10-007-262-1					
Query Match					
Best Local Similarity 100.0%; Score 2038; DB 9; Length 386;					
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MLPPKKKKLLFLVSGMAI	ALFFPHKYSNINISLSMKACPERMNVLVLSWRS	GSSFFVQ	60
DB	1	MLPPKKKKLLFLVSGMAI	ALFFPHKYSNINISLSMKACPERMNVLVLSWRS	GSSFFVQ	60
QY	61	LFQGHPIVFILMPAWVMTFKOSTAMLMHVPDILIRVFLCDMSVDPAYNEP	PPRQ	120	
DB	61	LFQGHPIVFILMPAWVMTFKOSTAMLMHVPDILIRVFLCDMSVDPAYNEP	PPRQ	120	
QY	121	SSLFQWNSALCSAPACDIIPODEIIPRAHCRLLCSQGFVEYERKACSYSHVLKEVR	180		
DB	121	SSLFQWNSALCSAPACDIIPODEIIPRAHCRLLCSQGFVEYERKACSYSHVLKEVR	180		
QY	181	FFNQSILYPLKPSINLHVLPDRPRAVRSRETKGMDISRTVWGQHQKXKED	240		
DB	181	FFNQSILYPLKPSINLHVLPDRPRAVRSRETKGMDISRTVWGQHQKXKED	240		











```

RESULT 13
US-09-927-602-11
; Sequence 11, Application US/09927602
; Patent No. US2002006156ZAI
GENERAL INFORMATION:
APPLICANT: FUKUDA, Michiko N.
APPLICANT: AKAMA, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LT 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR FILING DATE: 2000-06-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 174
TYPE: PR1
ORGANISM: Homo Sapien
US-09-927-602-11

Query Match      14.4% Score 293; DB 10; Length 174;
Best Local Similarity 25.6%; Pred. No. 7.9e-22;
Matches 69; Conservative 30; Mismatches 59; Indels 112; Gaps 2.

OY 42 RMHVLVSSRSGSSVGVGLFGOHDPVLYLKEPAHWVMYTFKOSTAWMLHMAYDLIRAV 101
| | | | | : | | | | | : | | | | | : | | | | |
DB 12 RRHVLNATRTGSSSVGGEEFNQGNITYLEPLMHIRTYS----- 53

OY 102 FLCDMSVFADYAMPEGRROSSLFCWENSRALCSAPACDIIPODEIIPRAHCRLCSQOPF 161
| | | | | : | | | | | : | | | | | : | | | | |
DB 54 ----- 53

OY 162 EYVKACRSYSHVYLKEVRFFNLQSILPKDPSLNTHIVLVDPRAVRRETRTKGDL 221
| | | | | : | | | | | : | | | | | : | | | | |
DB 54 --FEACRRKHMAKLAARIROLEFLQPLABPRIDLVIOLVNDPRAVLASR----- 104

OY 222 MIDSRIVMGHEKKLKEDQPIYVMQVICQSOLEIKTKITQSLPALOEIRULVRYEDLAR 281
| | | | | : | | | | | : | | | | | : | | | | |
DB 105 -----MVAFAGKYKTWELGLROPMALRGYRMVLVRYEEDVAR 139

OY 282 APVAQTSRMTEYVGLEFLPHLTGWVNINFR 311
| | | | | : | | | | | : | | | | |
DB 140 GPLQAKAREMYPFAGIPLTPQVEWMIDKNQ 169

RESULT 14
US-09-854-122-20
; Sequence 20, Application US/09854122
; Patent No. US20020016980A1
GENERAL INFORMATION:
APPLICANT: ALBERT, RANDALL S.
APPLICANT: SMITH, ROBERT
TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
FILE REFERENCE: PHA-007.01
CURRENT APPLICATION NUMBER: US/09/854,122
CURRENT FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/202,529
PRIOR FILING DATE: 2000-05-10
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 20
LENGTH: 320
TYPE: PR1
ORGANISM: Flaveria bidentis
US-09-854-122-20

Query Match      4.4% Score 89; DB 10; Length 320;
Best Local Similarity 20.8%; Pred. No. 0.39;
Matches 56; Conservative 50; Mismatches 83; Indels 100; Gaps 12
```

```
QY      54 GSSVGLFGQH--DVTYLMEPMHWMTGTOSTAMHLHAYVDLIRAVELDMISVFDAY 112
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      47 GALLAQGSFKRPDPVLYLCSP-----KSGTMYLKALAAVAIVR-----EKDFEF 91

QY      113 MEP-----GRRSSLFQ--WENSRALCSAPACDIIPDEIIFRAMCRLLSCOPE 162
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      92 TSPLLTIAPHNCIFYIKEDDKKIYENONNSCTP-----MÄTMPIH 133

QY      163 VVERAKRSYSNVLKEVFNFNLQSLPYLLKDPSLNLIHVLDPRVFRSRETRKGDM 222
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      134 VLPRS-----IL-----ALNCMIYYIYNINDV-----156

QY      223 IDSRITWGQHQRKKEDQPYVMQVTCQSOLEIYKITDSLP-----KALDER-- 270
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157 IVSYHFGRRETIKLPLDDAPE-----EAFDEFHGICSGFYRWHLIGWRASLEREE 210

QY      271 -YLLVRVEDLARAVAOSTRKYVEGLETF 298
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      211 VTLFLKIEDVKRDPTSNWKRALEIIGVPF 239
```

```

RESULT 15
US-09-902-941-1903
: Sequence 1903, Application US/09902941
: Patent No. US20020172952A1
: GENERAL INFORMATION:
: APPLICANT: Henderson, Robert A.
: APPLICANT: Wang, Tongtong
: APPLICANT: Watanabe, Yoshihiro
: APPLICANT: Johnson, Jeffrey C.
: APPLICANT: Retter, Marc W.
: APPLICANT: Marnerakis, Margarita
: APPLICANT: Carter, Derrick
: APPLICANT: Fanger, Gary R.
: APPLICANT: Vedavick, Thomas S.
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: McNabb, Andria
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.478C17
: CURRENT APPLICATION NUMBER: 05/09/902,941
: CURRENT FILING DATE: 2001-07-10
: NUMBER OF SEQ ID NOS: 2002
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1903
: LENGTH: 2209
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-902-941-1903

```

```

Query Match      4 3% Score 88; DB 9; Length 2209;
Best Local Similarity 16.2%; Pred. No. 7.6;
Matches 54; Conservative 48; Mismatches 96; Indels 134; Gaps

QY      29 HNISLSKKAQPERRHVLVLSMRSSSEFVGOLF--GHDPVFYL-----M 72
          :::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1253 NISLILNQNIGES-----SQWELALRFVWGSEFGCHLSSNMNNATLSEKLFEETLL 1305

QY      73 EPAMHWMTFKQSTAWMLLHAAVADLLRAYFLC---DMASFDAIMPEGRROSSLFQWMS 129
          : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1306 VKSRIVVAVELKEKAVIFIREENATLLHKYFNCELYDLDALCY----- 1348

QY      130 RALCSAPACDIIPDEDTIIPAHGCRILCSOOPEVEVKACSASHVLKEVFNFNLQSLV- 188
          : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1349 -----CTLLPKDFV-----EULMKTLDKAQNNTDKLIALSIWSSELASLIYO 1390

QY      189 -----PLIKDPSPNLHLVIHLVRDPAVFSRSPERTKGDL-----IDSKRIWMQ 231
          : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1391 ELEMGLKRELSTDQAQGIRLGKLGISFCQVEFHQLTKKKDLITAKALENIDMDTSLIL- 1448

QY      232 HEQRKKEKDQPIYVMQYICSOLEIFYKTYSLOSPKALQEQRYLVREDDLARNAVQDSIMY 291
          : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Mon Jan 13 09:14:06 2003

us-09-816-825-2.rapb

Page 7

Db 1449 -----EYCSPTQLDCDAVLQIF----- 1465

QY 292 EFVGLFPLPHJQFWVHNITRGKSGDHAFHJNAR 325

Db 1466 -----IETLHNTNAGQGQGDASMDSAKR 1489

Search completed: January 11, 2003, 01:22:57  
Job time : 22 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 21:40:22 ; Search time 5288 Seconds

(without alignments)  
11183.226 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032

Sequence: 1 ggcctcgagcagcagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 2054640 segs, 14551402878 residues

Word size: 20

Total number of hits satisfying chosen parameters: 1438

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_la:\*  
4: gb\_em:\*  
5: gb\_cv:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pla:\*  
35: em\_hgt\_rod:\*  
36: em\_hgt\_mam:\*  
37: em\_hgt\_vtl:\*  
38: em\_sy:\*  
39: em\_hgt\_hum:\*  
40: em\_hgt\_mus:\*  
41: em\_hgtg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2032	100.0	2032	6	AR203335	AR203335 Sequence
2	2032	100.0	2032	9	AF131225	AF131225 Homo sapi
3	1856	91.3	2011	9	AK026635	AK026635 Homo sapi
4	1806	88.9	183228	9	AC010547	AC010547 Homo sapi
5	1756	86.4	1992	9	AF280088	AF280088 Homo sapi
6	1211	59.6	1333	9	AF149783	AF149783 Homo sapi
7	505	24.9	517	6	AX381256	AX381256 Sequence
8	286	14.1	389	6	AX381326	AX381326 Sequence
9	259	2.9	1462	9	AF176839	AF176839 Homo sapi
10	59	2.9	1647	6	AX327330	AX327330 Sequence
11	59	2.9	2170	9	AF176838	AF176838 Homo sapi
12	59	2.9	2544	9	AF219990	AF219990 Homo sapi
13	59	2.9	3278	9	AF246718	AF246718 Homo sapi
14	59	2.9	3786	9	AF280086	AF280086 Homo sapi
15	59	2.9	71503	9	AF219991	AF219991 Homo sapi
16	59	2.9	157337	9	AC009163	AC009163 Homo sapi
17	59	2.9	194832	9	AC025287	AC025287 Homo sapi
18	59	2.9	208185	2	AC009105	AC009105 Homo sapi
19	37	1.8	37	6	AR203340	AR203340 Sequence
20	27	1.3	1926	10	AF131236	AF131236 Mus muscu
21	27	1.3	2201	10	AF109155	AF109155 Mus muscu
22	24	1.2	1398	9	AF302109	AF302109 Homo sapi
23	24	1.2	79388	2	AC012281	AC012281 Homo sapi
24	24	1.2	99395	2	AC095651	AC095651 Rattus no
25	24	1.2	131969	2	AC116944	AC116944 Tetradon
26	24	1.2	143655	2	AC010808	AC010808 Homo sapi
27	24	1.2	168991	9	AC073958	AC073958 Homo sapi
28	24	1.2	181557	2	AC072055	AC072055 Homo sapi
29	24	1.2	191639	2	AC116046	AC116046 Papio ham
30	23	1.1	1740	10	AF176841	AF176841 Mus muscu
31	23	1.1	1989	10	AF176840	AF176840 Mus muscu
32	23	1.1	2002	9	BC011365	BC011365 Homo sapi
33	23	1.1	2459	10	BC013482	BC013482 Mus muscu
34	23	1.1	2488	10	BC026628	BC026628 Mus muscu
35	23	1.1	2860	6	E12442	E12442 Dna encodir
36	23	1.1	29064	2	AC017237	AC017237 Drosophill
37	23	1.1	35470	2	AC116989	AC116989 Dictyoste
38	23	1.1	56342	2	AC103270	AC103270 Rattus no
39	23	1.1	63739	2	AC087621	AC087621 Homo sapi
40	23	1.1	69523	2	AC124305	AC124305 Homo sapi
41	23	1.1	74718	10	AC083858	AC083858 Mus muscu
42	23	1.1	83543	9	AC117427	AC117427 Homo sapi
43	23	1.1	92910	9	AL590368	AL590368 Human Dna
44	23	1.1	97847	9	HS874C20	HS874C20 Human Dna
45	23	1.1	109151	2	AC017637	AC017637 Drosophill
46	23	1.1	111669	2	AC128720	AC128720 Rattus no
47	23	1.1	133902	2	AC095664	AC095664 Rattus no
48	23	1.1	141017	2	AC116962	AC116962 Dictyoste
49	23	1.1	143351	2	AC099282	AC099282 Rattus no
50	23	1.1	159170	2	AC115914	AC115914 Mus muscu
51	23	1.1	160629	3	AC023715	AC023715 Drosophill
52	23	1.1	161278	3	AC008223	AC008223 Drosophill
53	23	1.1	164637	2	AC016192	AC016192 Homo sapi
54	23	1.1	172965	2	AC023367	AC023367 Homo sapi
55	23	1.1	172970	2	AC111112	AC111112 Mus muscu
56	23	1.1	177080	2	AC027288	AC027288 Homo sapi
57	23	1.1	177142	9	AC004582	AC004582 Human Chr
58	23	1.1	177357	2	AC124078	AC124078 Homo sapi
59	23	1.1	179618	2	AC068214	AC068214 Homo sapi
60	23	1.1	180169	9	AL590101	AL590101 Human Dna
61	23	1.1	180692	2	AC123320	AC123320 Rattus no
62	23	1.1	184591	2	AC117669	AC117669 Rattus no
63	23	1.1	185671	9	AC084193	AC084193 Homo sapi
64	23	1.1	187543	9	AC005678	AC005678 Homo sapi
65	23	1.1	190153	3	AC008201	AC008201 Drosophill

C	65	23	1.1	205429	2	AC005506	139	22	1.1	94234	2	AC112296	AC112296 Rattus no
C	67	23	1.1	215523	2	ALB84190	C 140	22	1.1	96433	2	DMB519	AL121800 Drosophila
C	68	23	1.1	217208	2	AC11927	C 141	22	1.1	97799	2	AL354890	AL354890 Homo sapi
C	69	23	1.1	237950	2	AC127315	C 142	22	1.1	97798	2	AC097051	AC097051 Rattus no
C	70	23	1.1	239327	3	AE003744	C 143	22	1.1	100726	2	AC116961	AC116961 Drosophila
C	71	23	1.1	288687	2	AC125315	C 144	22	1.1	102064	2	AC002558	AC002558 Homo sapi
C	72	23	1.1	319551	3	AE003432	C 145	22	1.1	104014	2	AC116921	AC116921 Drosophila
C	73	23	1.1	39252	3	DDACT15P	C 146	22	1.1	104483	2	AC106977	AC106977 Rattus no
C	74	22	1.1	707	3	AY075486	C 147	22	1.1	107609	2	AC118062	AC118062 Homo sapi
C	75	22	1.1	934	3	DD025660	C 148	22	1.1	110000	2	PEMALAP1	Continuation (2 of
C	76	22	1.1	1435	3	AF378371	C 149	22	1.1	110000	10	AE014177_1	AC078780 Homo sapi
C	77	22	1.1	1477	3	AY089472	C 150	22	1.1	110000	30	AC078780_0	AC078780 Homo sapi
C	78	22	1.1	1629	5	AF124511	C 151	22	1.1	111104	9	AC117389	AC117389 Homo sapi
C	79	22	1.1	1740	5	HUMEN51PRO	C 152	22	1.1	111173	3	AL513204	AL513204 Human DNA
C	80	22	1.1	1782	6	AX401786	C 153	22	1.1	113880	9	PEMAL3P4	AL008970 Plasmodium
C	81	22	1.1	1782	10	RNU32208	C 154	22	1.1	114748	2	AC103434	AC103434 Rattus no
C	82	22	1.1	1801	10	BC012744	C 155	22	1.1	114748	2	AC103434	AC103434 Rattus no
C	83	22	1.1	1801	9	AF289495	C 156	22	1.1	115018	9	AL162595	AL162595 Human DNA
C	84	22	1.1	2007	9	AF289495	C 157	22	1.1	115018	9	AL162595	AL162595 Human DNA
C	85	22	1.1	2115	9	AK024885	C 158	22	1.1	115545	2	AC119334	AC119334 Rattus no
C	86	22	1.1	2220	3	DDACT15	C 159	22	1.1	117283	2	CNS01RIL	AL163633 Human chr
C	87	22	1.1	2220	3	DDACT15	C 160	22	1.1	119696	9	HSDJ777L9	AL049794 Human DNA
C	88	22	1.1	2499	10	BC008135	C 161	22	1.1	121539	2	AC108240	AC108240 Rattus no
C	89	22	1.1	2558	3	PEBDAAS	C 162	22	1.1	123280	2	AC117076	AC117076 Drosophila
C	90	22	1.1	2610	3	HSB803995	C 163	22	1.1	125152	2	AC093936	AC093936 Rattus no
C	91	22	1.1	2678	8	AF360344	C 164	22	1.1	126143	2	AC098343	AC098343 Rattus no
C	92	22	1.1	2691	9	BC017210	C 165	22	1.1	131259	2	AP001805	AP001805 Homo sapi
C	93	22	1.1	2844	10	BC027428	C 166	22	1.1	134790	2	HSJ543J13	AL121999 Human DNA
C	94	22	1.1	2941	9	AK025582	C 167	22	1.1	137712	2	AC102683	AC102683 Mus muscu
C	95	22	1.1	3126	6	A21625	C 168	22	1.1	142232	8	AC102316	AC102316 Mus muscu
C	96	22	1.1	3357	6	AX277591	C 169	22	1.1	143248	2	AC107315	AC107315 Genomic s
C	97	22	1.1	3597	6	AX277591	C 170	22	1.1	143248	2	AC079855	AC079855 Homo sapi
C	98	22	1.1	4814	3	AY075427	C 171	22	1.1	145100	2	AC024641	AC024641 Homo sapi
C	99	22	1.1	4830	3	DME420377	C 172	22	1.1	146399	9	AC025835	AC025835 Homo sapi
C	100	22	1.1	6105	12	ASPDAX3H	C 173	22	1.1	146487	2	AC010803	AC010803 Homo sapi
C	101	22	1.1	6105	12	ASPDAX3H	C 174	22	1.1	147192	2	AC116925	AC116925 Drosophila
C	102	22	1.1	6113	12	ASPDAX3C	C 175	22	1.1	147750	2	AC110930	AC110930 Rattus no
C	103	22	1.1	6114	12	ASPDAX3C	C 176	22	1.1	148557	2	AC116827	AC116827 Mus muscu
C	104	22	1.1	6146	12	AF269235	C 177	22	1.1	151365	9	AC010382	AC010382 Homo sapi
C	105	22	1.1	6872	12	AF269235	C 178	22	1.1	151365	2	AC116591	AC116591 Mus muscu
C	106	22	1.1	8574	3	AF057019	C 179	22	1.1	151776	2	PEMAL1P1	AL031744 Plasmodium
C	107	22	1.1	10452	12	AF269238	C 180	22	1.1	152409	2	AC129073	AC129073 Fells cat
C	108	22	1.1	1165	12	AF269237	C 181	22	1.1	153360	2	AC095290	AC095290 Rattus no
C	109	22	1.1	12029	3	AE001385	C 182	22	1.1	154371	9	AP003466	AP003466 Homo sapi
C	110	22	1.1	12229	3	AE001377	C 183	22	1.1	155020	2	AC106390	AC106390 Rattus no
C	111	22	1.1	12940	3	AE001386	C 184	22	1.1	155191	2	AC119025	AC119025 Rattus no
C	112	22	1.1	17706	2	AC014224	C 185	22	1.1	155222	10	AL683881	AL683881 Mouse DNA
C	113	22	1.1	23676	2	AC115599	C 186	22	1.1	156060	2	AC004153	AC004153 Plasmodium
C	114	22	1.1	23919	2	AC091224	C 187	22	1.1	156506	2	AC068038	AC068038 Homo sapi
C	115	22	1.1	27314	9	AC053516	C 188	22	1.1	157069	2	AC123209	AC123209 Rattus no
C	116	22	1.1	36521	2	AC030203	C 189	22	1.1	157102	2	AC116409	AC116409 Homo sapi
C	117	22	1.1	37572	2	AC014052	C 190	22	1.1	157150	2	AL355813	AL355813 Human DNA
C	118	22	1.1	3765	2	AC109977	C 191	22	1.1	157336	2	AC095253	AC095253 Rattus no
C	119	22	1.1	44700	9	AL590825	C 192	22	1.1	157361	2	AC106572	AC106572 Rattus no
C	120	22	1.1	46879	2	AC015095	C 193	22	1.1	157976	2	AC127476	AC127476 Fells cat
C	121	22	1.1	53380	2	AL135792	C 194	22	1.1	158750	2	AC067981	AC067981 Homo sapi
C	122	22	1.1	54527	2	AC015174	C 195	22	1.1	158832	2	AC097950	AC097950 Rattus no
C	123	22	1.1	55038	2	AC015270	C 196	22	1.1	159774	2	AC099367	AC099367 Rattus no
C	124	22	1.1	55470	2	AC116968	C 197	22	1.1	160001	3	AC113619	AC113619 Drosophila
C	125	22	1.1	59866	2	AC100572	C 198	22	1.1	160330	2	AC025631	AC025631 Homo sapi
C	126	22	1.1	65201	3	AC004059	C 199	22	1.1	160759	2	AC117082	AC117082 Drosophila
C	127	22	1.1	67297	3	AC005135	C 200	22	1.1	161201	2	AC025673	AC025673 Homo sapi
C	128	22	1.1	68464	9	AC005248	C 201	22	1.1	161437	2	AC093434	AC093434 Homo sapi
C	129	22	1.1	69480	9	AC092587	C 202	22	1.1	163035	2	AC006405	AC006405 Homo sapi
C	130	22	1.1	69511	2	AC094686	C 203	22	1.1	163447	2	AC111523	AC111523 Rattus no
C	131	22	1.1	70578	2	AC100228	C 204	22	1.1	163879	2	AC099310	AC099310 Drosophila
C	132	22	1.1	73763	2	AC111965	C 205	22	1.1	164349	2	AC123886	AC123886 Homo sapi
C	133	22	1.1	77601	9	HS45P1	C 206	22	1.1	165732	2	AP004388	AP004388 Homo sapi
C	134	22	1.1	82469	8	AC116330	C 207	22	1.1	165829	2	AC118484	AC118484 Rattus no
C	135	22	1.1	83865	8	AB006700	C 208	22	1.1	167944	2	AL844157	AL844157 Mus muscu
C	136	22	1.1	91355	8	AL391417	C 209	22	1.1	168185	2	AC079805	AC079805 Homo sapi
C	137	22	1.1	93897	2	AC129831	C 210	22	1.1	168426	2	AC124357	AC124357 Mus muscu
C	138	22	1.1	93913	2	AC122576	C 211	22	1.1	168773	10	AL731820	AL731820 Mouse DNA



C 212	22	1.1	169161	2	AC120589	Rattus no	285	22	1.1	209463	9	CNS05TDV	AL357153	Human chr
C 213	22	1.1	169190	2	AC114776	Homo sapi	286	22	1.1	209636	2	AL732412	AL732412	Mus muscu
C 214	22	1.1	169308	2	AC124700	Mus muscu	C 287	22	1.1	210336	2	AC022171	AC022171	Homo sapi
C 215	22	1.1	169469	2	AC037195	Homo sapi	288	22	1.1	210631	2	AC129184	AC129184	Mus muscu
C 216	22	1.1	169510	2	AC129168	Rattus no	289	22	1.1	2113045	2	AC093640	AC093640	Homo sapi
C 217	22	1.1	169774	2	AC129622	Rattus no	290	22	1.1	216347	2	AL663095	AL663095	Mus muscu
C 218	22	1.1	169794	2	AC004688	Plasmodiu	291	22	1.1	216408	2	AC092466	AC092466	Homo sapi
C 219	22	1.1	171020	2	AC036225	Homo sapi	C 292	22	1.1	217164	2	AL844482	AL844482	Mus muscu
C 220	22	1.1	171286	3	AC008226	Drosophi1	C 293	22	1.1	217353	2	AC123403	AC123403	Mus muscu
C 221	22	1.1	171390	9	AC027315	Homo sapi	294	22	1.1	217409	6	AC123901	AC123901	Homo sapi
C 222	22	1.1	171471	2	AC113073	Mus muscu	C 295	22	1.1	217743	2	AC116293	AC116293	Rattus no
C 223	22	1.1	171873	2	AC024995	Homo sapi	296	22	1.1	222605	9	AC010973	AC010973	Homo sapi
C 224	22	1.1	172676	2	AC087736	Par trogl	297	22	1.1	223561	2	AC121574	AC121574	Mus muscu
C 225	22	1.1	172714	2	AL845489	Mus muscu	C 298	22	1.1	223879	9	AC008735	AC008735	Homo sapi
C 226	22	1.1	173168	3	AC093196	Drosophi1	C 299	22	1.1	224448	2	PEMAL14P4	PEMAL14P4	Mus muscu
C 227	22	1.1	174323	2	AC092754	Homo sapi	300	22	1.1	224840	2	AC027284	AC027284	Mus muscu
C 228	22	1.1	174378	2	AC020983	Homo sapi	301	22	1.1	226013	2	AC073680	AC073680	Mus muscu
C 229	22	1.1	174579	2	AC016666	Homo sapi	C 302	22	1.1	226013	2	AC073680	AC073680	Mus muscu
C 230	22	1.1	174594	2	AC098986	Rattus no	303	22	1.1	227506	9	AC079969	AC079969	Homo sapi
C 231	22	1.1	175856	2	AC102209	Mus muscu	304	22	1.1	229045	2	AC008680	AC008680	Homo sapi
C 232	22	1.1	176095	2	AC019346	Homo sapi	305	22	1.1	230110	2	AC122396	AC122396	Mus muscu
C 233	22	1.1	177203	2	AC055784	Homo sapi	C 306	22	1.1	233573	3	AC091340	AC091340	Rattus no
C 234	22	1.1	178626	2	AC110511	Mus muscu	C 307	22	1.1	234651	2	AE0033740	AE0033740	Drosophi1
C 235	22	1.1	179039	2	AC102295	Mus muscu	308	22	1.1	236246	2	AC099068	AC099068	Rattus no
C 236	22	1.1	179396	3	AC022351	Drosophi1	C 309	22	1.1	238554	2	AC020965	AC020965	Mus muscu
C 237	22	1.1	179623	3	AC015429	Homo sapi	310	22	1.1	245802	2	AC006279	AC006279	Plasmodiu
C 238	22	1.1	179795	3	AC105351	AP003630	C 311	22	1.1	256861	2	AC129078	AC129078	Mus muscu
C 239	22	1.1	179902	2	AP003630	Oryza sat	C 312	22	1.1	265236	2	AE003626	AE003626	Drosophi1
C 240	22	1.1	180790	2	AC106172	Rattus no	C 313	22	1.1	268147	2	AC116966	AC116966	Dictyoste
C 241	22	1.1	181039	2	AC106915	Rattus no	C 314	22	1.1	268459	2	AC098250	AC098250	Rattus no
C 242	22	1.1	181250	9	AC090403	Homo sapi	315	22	1.1	292420	2	AC020873	AC020873	Mus muscu
C 243	22	1.1	181328	2	AC104945	Homo sapi	C 316	22	1.1	293431	2	PEMAL13P4	PEMAL13P4	Plasmodiu
C 244	22	1.1	182366	2	AC041041	Homo sapi	317	22	1.1	301980	3	AE003502	AE003502	Drosophi1
C 245	22	1.1	182600	2	AC034126	AC090567	C 318	22	1.1	306135	3	AE003436	AE003436	Drosophi1
C 246	22	1.1	182622	2	AC090567	Homo sapi	319	22	1.1	309023	3	AE003671	AE003671	Drosophi1
C 247	22	1.1	182660	2	AC012416	AC009825	320	22	1.1	310779	2	AC005140	AC005140	Plasmodiu
C 248	22	1.1	183989	2	AC009825	Homo sapi	321	22	1.1	310909	2	AE003437	AE003437	Drosophi1
C 249	22	1.1	184026	2	AC022176	Homo sapi	322	22	1.1	316084	2	AC127345	AC127345	Mus muscu
C 250	22	1.1	185937	2	AC094540	Rattus no	323	22	1.1	321003	9	PEMAL1P3	PEMAL1P3	Plasmodiu
C 251	22	1.1	186092	9	AC020556	Homo sapi	C 324	22	1.1	195	2	HS79D1R	HS79D1R	H.sapiens C
C 252	22	1.1	186323	9	AC009405	Homo sapi	C 325	22	1.1	196	9	HS47F4R	HS47F4R	H.sapiens C
C 253	22	1.1	186559	10	AL606742	Mouse DNA	326	22	1.1	196	9	HS79D1R	HS79D1R	H.sapiens C
C 254	22	1.1	187730	2	AC110613	Homo sapi	327	22	1.1	294	6	AF082830	AF082830	Taenia so
C 255	22	1.1	188246	9	AC013565	Homo sapi	328	22	1.1	284	6	AX083750	AX083750	Sequence
C 256	22	1.1	188459	2	AC008312	Drosophi1	329	22	1.1	329	3	AF082829	AF082829	Taenia so
C 257	22	1.1	188835	2	AC104155	Papio cyn	330	22	1.1	329	3	AF257776	AF257776	Taenia so
C 258	22	1.1	189008	2	AC125917	Rattus no	331	22	1.1	364	3	AF163972	AF163972	Taenia so
C 259	22	1.1	1900727	9	AC068724	Homo sapi	332	22	1.1	372	2	AF216695	AF216695	Taenia so
C 260	22	1.1	190911	2	AC095406	Rattus no	C 333	22	1.1	402	8	AF427562	AF427562	Bikinia a
C 261	22	1.1	191093	10	AL671765	Mouse DNA	C 334	22	1.1	410	6	AF427568	AF427568	Bikinia g
C 262	22	1.1	191312	2	AP001030	Homo sapi	C 335	22	1.1	411	6	AX330823	AX330823	Sequence
C 263	22	1.1	191750	2	AC122498	Mus muscu	C 336	22	1.1	411	6	AX408781	AX408781	Sequence
C 264	22	1.1	193038	2	AC128823	Rattus no	337	22	1.1	415	3	AB044083	AB044083	Taenia so
C 265	22	1.1	193674	2	AC125111	Mus muscu	C 338	22	1.1	437	8	AF427552	AF427552	Tetraheryl
C 266	22	1.1	194000	2	AC000016	Homo sapi	C 339	22	1.1	439	9	AF427578	AF427578	Aphanocai
C 267	22	1.1	194286	2	AC129160	Rattus no	340	22	1.1	439	9	BC031337	BC031337	Homo sapi
C 268	22	1.1	194531	2	AC124924	Rattus no	C 341	22	1.1	455	8	AF427581	AF427581	Bikinia p
C 269	22	1.1	194552	2	AC1083327	Rattus no	C 342	22	1.1	456	8	AF427548	AF427548	Brechyste
C 270	22	1.1	194633	2	AC090362	Homo sapi	C 343	22	1.1	457	8	AF427383	AF427383	Bikinia b
C 271	22	1.1	195460	2	AL672289	Homo sapi	C 344	22	1.1	458	8	AF427557	AF427557	Julbernar
C 272	22	1.1	195727	2	AC036234	Homo sapi	C 345	22	1.1	458	8	AF427566	AF427566	Bikinia d
C 273	22	1.1	195744	9	ALC092002	Homo sapi	C 346	22	1.1	465	8	AF427575	AF427575	Aphanocai
C 274	22	1.1	196000	2	AL354749	Human DNA	C 347	22	1.1	465	8	AF427580	AF427580	Bikinia m
C 275	22	1.1	198124	2	AC111459	Rattus no	C 348	22	1.1	466	8	AF427544	AF427544	Aphanocai
C 276	22	1.1	198487	2	AC128485	Rattus no	C 349	22	1.1	467	8	AF427579	AF427579	Aphanocai
C 277	22	1.1	199551	2	AC006281	Plasmodiu	C 350	22	1.1	477	8	AF427577	AF427577	Aphanocai
C 278	22	1.1	203904	2	AP002077	Homo sapi	C 351	22	1.1	484	8	AF427559	AF427559	Aphanocai
C 279	22	1.1	203493	2	AC058803	Homo sapi	352	22	1.1	600	9	BC020826	BC020826	Homo sapi
C 280	22	1.1	203924	2	AC012159	AC012159	353	22	1.1	660	9	BC006813	BC006813	Mus muscu
C 281	22	1.1	204652	2	PEMAL13P6	Plasmodiu	354	22	1.1	663	10	AY113211	AY113211	Drosophi1
C 282	22	1.1	205429	2	AC005506	Plasmodiu	355	22	1.1	764	3	BC002095	BC002095	Mus muscu
C 283	22	1.1	206055	2	AL844208	Mus muscu	356	22	1.1	808	10	AF471666	AF471666	illex main
C 284	22	1.1	208671	2	AC126498	Rattus no	C 357	22	1.1	853	8	AF471666	AF471666	illex main

358	21	1.0	873	11	CNS06E2J	AL395873 73 end of	c 431	21	1.0	93735	8	ATAC008261	AC008261 Arabidops
359	21	1.0	928	3	BC029527	BC029527 Homo sapi	c 432	21	1.0	95560	2	AC103136	AC103136 Rattus no
360	21	1.0	1042	3	AY070627	AY070627 Drosophil	c 433	21	1.0	98062	2	RN516010	AL003713 Rattus no
361	21	1.0	1101	10	BC022557	BC022557 Mus muscu	c 434	21	1.0	99725	8	ATF18674	AL021637 Arabidops
362	21	1.0	1165	8	AY065280	AY065280 Arabidops	c 435	21	1.0	99920	2	AC096218	AC096218 Rattus no
363	21	1.0	1280	8	AY061225	AY061225 Drosophil	c 436	21	1.0	100409	2	AC103320	AC103320 Rattus no
364	21	1.0	1282	10	MM0238978	AJ238978 Mus muscu	c 437	21	1.0	100511	8	AC010774	AC010774 Homo sapi
365	21	1.0	1364	9	BC030144	BC030144 Homo sapi	c 438	21	1.0	100585	2	AC023754	AC023754 Arabidops
366	21	1.0	1439	10	BC005457	BC005457 Mus muscu	c 439	21	1.0	102241	8	AC097198	AC097198 Rattus no
367	21	1.0	1440	8	TAFSBDCL	X07742 wheat chlor	c 440	21	1.0	104001	8	AC023528	AC023528 Arabidops
368	21	1.0	1469	6	AX375585	AX375585 Sequence	c 441	21	1.0	104163	8	AC010155	AC010155 Genomic s
369	21	1.0	1486	6	BC007016	BC007016 Homo sapi	c 442	21	1.0	105507	2	AC109113	AC109113 Rattus no
370	21	1.0	1547	8	AF160729	AF160729 Arabidops	c 443	21	1.0	106490	2	AC095569	AC095569 Rattus no
371	21	1.0	1593	9	BC008841	BC008841 Homo sapi	c 444	21	1.0	110000	2	AC095560-1	Continuation (2 of
372	21	1.0	1594	8	CEY17297	Y17297 Cunninghame	c 445	21	1.0	111682	2	AC005162	AC005162 Homo sapi
373	21	1.0	1746	8	BC011614	BC011614 Homo sapi	c 446	21	1.0	115028	2	AC099177	AC099177 Rattus no
374	21	1.0	1793	8	AY070432	AY070432 Arabidops	c 447	21	1.0	118426	2	HS431C21	AC099177 Rattus no
375	21	1.0	1967	3	AY060688	AY060688 Drosophil	c 448	21	1.0	119035	2	AC096954	AC096954 Rattus no
376	21	1.0	2041	3	ALRDNA1	X05836 Ascaris lum	c 449	21	1.0	119063	2	AC027311	AC027311 Homo sapi
377	21	1.0	2145	3	AY070598	AY070598 Drosophil	c 450	21	1.0	123039	2	AC116268	AC116268 Rattus no
378	21	1.0	2146	3	AY113401	AY113401 Drosophil	c 451	21	1.0	125144	2	AC099359	AC099359 Rattus no
379	21	1.0	2151	10	BC005571	BC005571 Mus muscu	c 452	21	1.0	129218	2	AC093380	AC093380 Homo sapi
380	21	1.0	2177	9	AF236871	AF236871 Homo sapi	c 453	21	1.0	129719	9	AL358113	AL358113 Human DNA
381	21	1.0	2230	8	CM2450	X76226 Candida mal	c 454	21	1.0	130834	2	AC115231	AC115231 Rattus no
382	21	1.0	2262	3	AF079448	AF079448 Dictyoste	c 455	21	1.0	130923	2	AC116295	AC116295 Rattus no
383	21	1.0	2403	5	AF014370	AF014370 Danio rer	c 456	21	1.0	131245	2	AC113320	AC113320 Human DNA
384	21	1.0	2423	9	BC009981	BC009981 Homo sapi	c 457	21	1.0	132596	2	AC013046	AC013046 Drosophil
385	21	1.0	2912	9	AF261655	AF261655 Homo sapi	c 458	21	1.0	132966	2	AC125403	AC125403 Mus muscu
386	21	1.0	3160	10	AF102887	AF102887 Mus muscu	c 459	21	1.0	133255	2	AC121410	AC121410 Rattus no
387	21	1.0	3379	3	AF111943	AF111943 Dictyoste	c 460	21	1.0	133276	2	AC087147	AC087147 Mus muscu
388	21	1.0	3391	10	F32889S11	AF328994 Mus muscu	c 461	21	1.0	133501	2	AC116556	AC116556 Dictyoste
389	21	1.0	3479	3	NEMRGA	M74384 Ascaris lum	c 462	21	1.0	134190	2	AC128770	AC128770 Rattus no
390	21	1.0	3597	3	AY118761	AY118761 Drosophil	c 463	21	1.0	134372	2	AC118807	AC118807 Rattus no
391	21	1.0	4110	3	PFASRRP	M55428 P.falciparu	c 464	21	1.0	134545	8	AB042240	AB042240 Trilicium
392	21	1.0	4155	8	CMCHMGCOAR	L10390 Camptotheca	c 465	21	1.0	137154	2	AL513082	AL513082 Human DNA
393	21	1.0	4226	10	BC019138	BC019138 Mus muscu	c 466	21	1.0	138019	2	AC121737	AC121737 Rattus no
394	21	1.0	4284	3	DDU02651	U20661 Dictyoste11	c 467	21	1.0	138859	2	AL359076	AL359076 Human DNA
395	21	1.0	4638	9	HSN803395	AL831998 Homo sapi	c 468	21	1.0	139390	2	AC110432	AC110432 Rattus no
396	21	1.0	4669	10	AF029240	AF029240 Rattus no	c 469	21	1.0	139555	2	AC025636	AC025636 Homo sapi
397	21	1.0	5310	2	AC012823	AC012823 Drosophil	c 470	21	1.0	140714	2	RN574E16	AL003712 Rattus no
398	21	1.0	6608	6	AX346936	AE465306 Dictyoste	c 471	21	1.0	143244	9	AC091997	AC091997 Homo sapi
399	21	1.0	9098	3	AE465306	AE465306 Dictyoste	c 472	21	1.0	144476	2	AC117283	AC117283 Rattus no
400	21	1.0	9671	1	AE006059	AE006059 Pasteurel	c 473	21	1.0	144542	2	AC015819	AC015819 Homo sapi
401	21	1.0	12029	3	AE001427	AE001427 Dictyoste	c 474	21	1.0	145124	2	AC022894	AC022894 Homo sapi
402	21	1.0	12355	2	AC115605	AC115605 Dictyoste	c 475	21	1.0	145577	2	AC114233	AC114233 Rattus no
403	21	1.0	12593	3	AE001418	AE001418 Plasmodiu	c 476	21	1.0	147775	9	AC015656	AC015656 Homo sapi
404	21	1.0	12919	2	AC115580	AC115580 Dictyoste	c 477	21	1.0	147882	2	AC118083	AC118083 Rattus no
405	21	1.0	14760	3	AE001410	AE001410 Plasmodiu	c 478	21	1.0	149599	2	AC112375	AC112375 Rattus no
406	21	1.0	15619	2	AL590650	AL590650 Human DNA	c 479	21	1.0	151455	2	AC095571	AC095571 Rattus no
407	21	1.0	18010	2	AC115607	AC115607 Dictyoste	c 480	21	1.0	152116	2	AC125089	AC125089 Mus muscu
408	21	1.0	20043	2	AC116986	AC116986 Dictyoste	c 481	21	1.0	152665	2	AC027005	AC027005 Homo sapi
409	21	1.0	33404	2	AC115684	AC115684 Dictyoste	c 482	21	1.0	152717	2	AC119505	AC119505 Rattus no
410	21	1.0	38872	2	AC068113	AC068113 Homo sapi	c 483	21	1.0	153004	9	AC096754	AC096754 Homo sapi
411	21	1.0	43324	2	AC034287	AC034287 Mus muscu	c 484	21	1.0	153798	2	AC108343	AC108343 Rattus no
412	21	1.0	48699	2	AC115677	AC115677 Dictyoste	c 485	21	1.0	153993	2	AC012262	AC012262 Homo sapi
413	21	1.0	56099	2	AC115598	AC115598 Dictyoste	c 486	21	1.0	154076	2	AC023273	AC023273 Homo sapi
414	21	1.0	58879	2	AC090350	AC090350 Homo sapi	c 487	21	1.0	154084	9	AC092034	AC092034 Homo sapi
415	21	1.0	63117	2	AC116505	AC116505 Mus muscu	c 488	21	1.0	154222	2	AC130631	AC130631 Rattus no
416	21	1.0	67757	9	AL772388	AL772388 Human DNA	c 489	21	1.0	154329	2	DM8925B3	AL398972 Drosophil
417	21	1.0	69578	2	AC118025	AC118025 Mus muscu	c 490	21	1.0	154348	9	AL691514	AL691514 Human DNA
418	21	1.0	71319	2	AC009670	AC009670 Homo sapi	c 491	21	1.0	154383	2	AC113760	AC113760 Rattus no
419	21	1.0	72541	9	AC007022	AC007022 Homo sapi	c 492	21	1.0	155584	2	AL161936	AL161936 Human DNA
420	21	1.0	75852	9	AC074133	AC074133 Homo sapi	c 493	21	1.0	155606	2	AC105664	AC105664 Rattus no
421	21	1.0	76578	2	AC109578	AC109578 Rattus no	c 494	21	1.0	155690	2	AC096042	AC096042 Rattus no
422	21	1.0	81766	2	AC112134	AC112134 Homo sapi	c 495	21	1.0	156358	2	AC012298	AC012298 Homo sapi
423	21	1.0	82893	8	AB022218	AB022218 Arabidops	c 496	21	1.0	156599	2	AC130743	AC130743 Rattus no
424	21	1.0	82893	8	AC024081	AC024081 Arabidops	c 497	21	1.0	157081	2	AC023084	AC023084 Homo sapi
425	21	1.0	84194	8	AB006701	AB006701 Arabidops	c 498	21	1.0	157097	2	AC024295	AC024295 Homo sapi
426	21	1.0	84648	2	AC016461	AC016461 Homo sapi	c 499	21	1.0	157677	2	AP002470	AP002470 Homo sapi
427	21	1.0	86106	8	AC011282	AC011282 Homo sapi	c 500	21	1.0	158055	2	AC127689	AC127689 Rattus no
428	21	1.0	88306	8	NCR7K22	AL670543 Neurospor	c 501	21	1.0	158554	2	AC112550	AC112550 Rattus no
429	21	1.0	92906	2	AC096459	AC096459 Rattus no	c 502	21	1.0	158558	2	AC094305	AC094305 Rattus no
430	21	1.0	93581	9	HS1059H15	AL022100 Human DNA	c 503	21	1.0	158615	2	AC117835	AC117835 Rattus no

504	21	1.0	159505	2	AC023142	577	21	1.0	189981	2	AC061986	AC061986 Homo sapi
505	21	1.0	160673	10	AL713977	578	21	1.0	190379	2	AC062386	AC062386 Homo sapi
506	21	1.0	161593	2	AC114387	579	21	1.0	190511	2	AC118096	AC118096 Rattus no
507	21	1.0	161813	9	CNS01RIG	580	21	1.0	190583	2	AC124574	AC124574 Mus muscu
508	21	1.0	163808	9	AC015961	581	21	1.0	190788	2	AC120812	AC120812 Rattus no
509	21	1.0	164007	9	AC095155	582	21	1.0	191161	2	AC107539	AC107539 Rattus no
510	21	1.0	164018	9	AL354868	583	21	1.0	191191	2	HSAC000119	AC000119 Human BAC
511	21	1.0	164423	2	AC026828	584	21	1.0	191342	3	AC104140	AC104140 Drosophill
512	21	1.0	164819	9	AC092335	585	21	1.0	191354	2	AC102648	AC102648 Mus muscu
513	21	1.0	164970	2	AC098136	586	21	1.0	191553	2	AC094883	AC094883 Rattus no
514	21	1.0	165416	9	AC034103	587	21	1.0	192003	10	AC084292	AC084292 Mus muscu
515	21	1.0	165970	2	AC098052	588	21	1.0	192087	2	AC079325	AC079325 Homo sapi
516	21	1.0	166341	2	AL513486	589	21	1.0	192188	2	AC109381	AC109381 Rattus no
517	21	1.0	166503	2	AC023577	590	21	1.0	192367	2	AC107393	AC107393 Homo sapi
518	21	1.0	166534	2	AC034209	591	21	1.0	192517	2	AC094413	AC094413 Rattus no
519	21	1.0	166692	2	AC023475	592	21	1.0	192959	2	AC117065	AC117065 Rattus no
520	21	1.0	166839	2	AC130987	593	21	1.0	193188	2	AC067728	AC067728 Homo sapi
521	21	1.0	167694	9	AC113137	594	21	1.0	193459	2	AC097966	AC097966 Rattus no
522	21	1.0	168154	9	AC114098	595	21	1.0	193659	10	AC122438	AC122438 Mus muscu
523	21	1.0	168168	2	AC012283	596	21	1.0	194509	2	AC107234	AC107234 Mus muscu
524	21	1.0	168954	2	AC099706	597	21	1.0	194860	2	AC113940	AC113940 Mus muscu
525	21	1.0	169101	2	AC008317	598	21	1.0	194994	9	AC108650	AC108650 Homo sapi
526	21	1.0	169311	3	AC123805	599	21	1.0	194995	10	AL627103	AL627103 Mouse DNA
527	21	1.0	169354	2	AC099598	600	21	1.0	196216	2	AC019228	AC019228 Homo sapi
528	21	1.0	170064	2	AC095307	601	21	1.0	196235	9	AC109810	AC109810 Homo sapi
529	21	1.0	170347	3	AC104054	602	21	1.0	196828	2	AC127709	AC127709 Rattus no
530	21	1.0	170443	9	AC0110863	603	21	1.0	197345	2	AC110248	AC110248 Mus muscu
531	21	1.0	170491	2	AC019105	604	21	1.0	197631	2	AC098511	AC098511 Rattus no
532	21	1.0	170711	2	AC032020	605	21	1.0	197797	2	AC012514	AC012514 Homo sapi
533	21	1.0	170736	2	AC013658	606	21	1.0	198427	8	ATCHRIV52	AL161552 Arabidops
534	21	1.0	171272	2	AC116684	607	21	1.0	199031	2	AC110416	AC110416 Rattus no
535	21	1.0	171452	9	AP000476	608	21	1.0	199052	2	AC122962	AC122962 Rattus no
536	21	1.0	172191	2	AL845505	609	21	1.0	199103	2	AC123813	AC123813 Mus muscu
537	21	1.0	172649	2	AC109744	610	21	1.0	199254	2	AC021144	AC021144 Homo sapi
538	21	1.0	173142	2	AC091897	611	21	1.0	199484	2	AC129057	AC129057 Rattus no
539	21	1.0	173295	9	AC099744	612	21	1.0	199579	2	AC117899	AC117899 Rattus no
540	21	1.0	173398	2	AC112985	613	21	1.0	199599	9	AC093536	AC093536 Homo sapi
541	21	1.0	173349	2	AC106504	614	21	1.0	201724	2	AC119506	AC119506 Rattus no
542	21	1.0	174678	9	AL450332	615	21	1.0	201879	2	AC122011	AC122011 Mus muscu
543	21	1.0	176036	3	AC009258	616	21	1.0	203713	2	AC084725	AC084725 Mus muscu
544	21	1.0	177104	10	AC091237	617	21	1.0	204192	2	AC121396	AC121396 Rattus no
545	21	1.0	177404	2	AC108296	618	21	1.0	204340	2	AC019238	AC019238 Homo sapi
546	21	1.0	178245	2	AC107343	619	21	1.0	204959	2	AC103945	AC103945 Mus muscu
547	21	1.0	178273	2	AC005308	620	21	1.0	208405	2	AC079477	AC079477 Mus muscu
548	21	1.0	178452	2	AC024672	621	21	1.0	209815	2	AL845290	AL845290 Mus muscu
549	21	1.0	178561	2	AC097685	622	21	1.0	211542	2	AL672182	AL672182 Mus muscu
550	21	1.0	178756	9	AC104420	623	21	1.0	211571	2	AC125765	AC125765 Rattus no
551	21	1.0	178942	2	AL357520	624	21	1.0	211871	2	AL845293	AL845293 Mus muscu
552	21	1.0	180340	9	AC016233	625	21	1.0	213216	9	HS399M15	AL449663 Homo sapi
553	21	1.0	180340	2	AC016233	626	21	1.0	214267	2	AL844881	AL844881 Mus muscu
554	21	1.0	180572	2	AC113894	627	21	1.0	214737	2	AC094443	AC094443 Rattus no
555	21	1.0	181365	2	AC131194	628	21	1.0	215267	2	AC079432	AC079432 Mus muscu
556	21	1.0	181712	2	AP001103	629	21	1.0	216120	2	AC021477	AC021477 Mus muscu
557	21	1.0	182052	2	AC119476	630	21	1.0	216185	2	AL833780	AL833780 Mus muscu
558	21	1.0	182314	2	AC013670	631	21	1.0	216671	2	AC126597	AC126597 Rattus no
559	21	1.0	183022	2	AC109103	632	21	1.0	217634	2	AC121979	AC121979 Mus muscu
560	21	1.0	183099	2	AC025288	633	21	1.0	219491	2	AC123057	AC123057 Homo sapi
561	21	1.0	183658	2	AC032021	634	21	1.0	220159	2	AC124507	AC124507 Mus muscu
562	21	1.0	183690	9	AC091180	635	21	1.0	221669	2	AC113026	AC113026 Mus muscu
563	21	1.0	184013	2	AC112632	636	21	1.0	222281	2	AC096335	AC096335 Rattus no
564	21	1.0	184284	2	AC025582	637	21	1.0	222826	2	AC124728	AC124728 Mus muscu
565	21	1.0	184531	2	AC112677	638	21	1.0	224788	9	AC011501	AC011501 Homo sapi
566	21	1.0	185108	2	AC118835	639	21	1.0	226803	2	AC123057	AC123057 Mus muscu
567	21	1.0	187064	9	AC011476	640	21	1.0	227813	2	AL772225	AL772225 Mus muscu
568	21	1.0	187505	2	AC121401	641	21	1.0	229037	2	AC097325	AC097325 Homo sapi
569	21	1.0	187561	2	AC130850	642	21	1.0	232278	2	AC087876	AC087876 Mus muscu
570	21	1.0	187921	30	AC096395	643	21	1.0	233129	2	AC094132	AC094132 Rattus no
571	21	1.0	188129	2	AC095782	644	21	1.0	233775	2	AC129336	AC129336 Mus muscu
572	21	1.0	188143	2	AL671872	645	21	1.0	234960	2	AC115675	AC115675 Rattus no
573	21	1.0	188260	2	AC110730	646	21	1.0	235571	2	AC128364	AC128364 Rattus no
574	21	1.0	188644	9	AC090518	647	21	1.0	235948	2	AC099574	AC099574 Mus muscu
575	21	1.0	188818	2	AC114019	648	21	1.0	236697	2	AC096364	AC096364 Rattus no
576	21	1.0	189340	2	AC051647	649	21	1.0	238528	2	AC098896	AC098896 Rattus no

c 650	21	1.0	238593	2	AC093343	Mus muscu	723	20	1.0	1354	9	BC001284	BC001284 Homo sapi
c 651	21	1.0	239927	2	AL731735	Mus muscu	724	20	1.0	1360	8	ATM24640	ATM24640 Arabidops
c 652	21	1.0	240809	2	AC096352	Rattus no	725	20	1.0	1366	8	AP239740	AP239740 Vitis vin
c 653	21	1.0	242677	2	AC113257	Rattus no	726	20	1.0	1366	8	BC020889	BC020889 Homo sapi
c 654	21	1.0	242677	2	AC113257	Rattus no	727	20	1.0	1436	8	BC020889	BC020889 Homo sapi
c 655	21	1.0	245383	2	AL161653	Homo sapi	728	20	1.0	1446	8	BC009205	BC009205 Arabidops
c 656	21	1.0	250925	2	AC123053	Mus muscu	729	20	1.0	1451	8	AC093053	AC093053 Homo sapi
c 657	21	1.0	251202	2	AC091902	Mus muscu	730	20	1.0	1463	3	AY051872	AY051872 Drosophill
c 658	21	1.0	253305	3	PFMAL137	Homo sapi	731	20	1.0	1469	8	BC028026	BC028026 Homo sapi
c 659	21	1.0	254993	2	AC078865	Mus muscu	732	20	1.0	1504	8	AY087356	AY087356 Arabidops
c 660	21	1.0	256172	2	AC005139	Plasmodiu	733	20	1.0	1526	9	AF131852	AF131852 Homo sapi
c 661	21	1.0	258024	2	AC084046	Typanoso	734	20	1.0	1548	3	AF068262	AF068262 Drosophill
c 662	21	1.0	258456	2	AC118802	Rattus no	735	20	1.0	1587	3	AF067206	AF067206 Drosophill
c 663	21	1.0	289090	3	AE003424	AE003424 Drosophill	736	20	1.0	1590	3	AF065473	AF065473 Drosophill
c 664	21	1.0	293431	2	PFMAL13B4	AL049181 Plasmodiu	737	20	1.0	1634	8	AY007108	AY007108 Homo sapi
c 665	21	1.0	310001	9	AF178030	Homo sapi	738	20	1.0	1649	8	HBRI32581	HBRI32581 Homo sapi
c 666	21	1.0	310779	2	AC005140	Plasmodiu	739	20	1.0	1651	8	HBRI32580	HBRI32580 Homo sapi
c 667	21	1.0	318221	2	PFMAL13B3	AL049184 Plasmodiu	740	20	1.0	1653	8	BHSAMC16	BHSAMC16 Homo sapi
c 668	21	1.0	329362	3	AE003681	AE003681 Drosophill	741	20	1.0	1725	9	BC034430	BC034430 Homo sapi
c 669	21	1.0	340000	9	AP001690	AP001690 Homo sapi	742	20	1.0	1734	9	BC016868	BC016868 Homo sapi
c 670	20	1.0	340000	9	AP001690	AP001690 Homo sapi	743	20	1.0	1737	9	HSMB00182	AL049397 Homo sapi
c 671	20	1.0	340000	9	AP001690	AP001690 Homo sapi	744	20	1.0	1750	10	BC018512	BC018512 Mus muscu
c 672	20	1.0	340000	9	AP001690	AP001690 Homo sapi	745	20	1.0	1833	10	BC004687	BC004687 Mus muscu
c 673	20	1.0	340000	9	AP001690	AP001690 Homo sapi	746	20	1.0	1855	9	AB060855	AB060855 Macaca fa
c 674	20	1.0	340000	9	AP001690	AP001690 Homo sapi	747	20	1.0	1869	8	ATSKIN2	ATSKIN2 Homo sapi
c 675	20	1.0	340000	9	AP001690	AP001690 Homo sapi	748	20	1.0	1913	9	AF053369	AF053369 Homo sapi
c 676	20	1.0	340000	9	AP001690	AP001690 Homo sapi	749	20	1.0	1956	8	AF036328	AF036328 Arabidops
c 677	20	1.0	340000	9	AP001690	AP001690 Homo sapi	750	20	1.0	1981	8	HSMB02144	AL137654 Homo sapi
c 678	20	1.0	340000	9	AP001690	AP001690 Homo sapi	751	20	1.0	2026	3	AE071409	AE071409 Plasmodiu
c 679	20	1.0	340000	9	AP001690	AP001690 Homo sapi	752	20	1.0	2051	9	AK000527	AK000527 Homo sapi
c 680	20	1.0	340000	9	AP001690	AP001690 Homo sapi	753	20	1.0	2128	3	AF027825	AF027825 Plasmodiu
c 681	20	1.0	340000	9	AP001690	AP001690 Homo sapi	754	20	1.0	2134	9	AK096948	AK096948 Homo sapi
c 682	20	1.0	340000	9	AP001690	AP001690 Homo sapi	755	20	1.0	2136	9	AK025213	AK025213 Homo sapi
c 683	20	1.0	340000	9	AP001690	AP001690 Homo sapi	756	20	1.0	2209	8	AY093170	AY093170 Arabidops
c 684	20	1.0	340000	9	AP001690	AP001690 Homo sapi	757	20	1.0	2228	10	BC025940	BC025940 Mus muscu
c 685	20	1.0	340000	9	AP001690	AP001690 Homo sapi	758	20	1.0	2256	10	BC006871	BC006871 Mus muscu
c 686	20	1.0	340000	9	AP001690	AP001690 Homo sapi	759	20	1.0	2271	5	AF014367	AF014367 Homo sapi
c 687	20	1.0	340000	9	AP001690	AP001690 Homo sapi	760	20	1.0	2320	9	AF139207	AF139207 Arabidops
c 688	20	1.0	340000	9	AP001690	AP001690 Homo sapi	761	20	1.0	2407	9	BC018361	BC018361 Homo sapi
c 689	20	1.0	340000	9	AP001690	AP001690 Homo sapi	762	20	1.0	2429	10	BC011270	BC011270 Mus muscu
c 690	20	1.0	340000	9	AP001690	AP001690 Homo sapi	763	20	1.0	2564	8	AY034972	AY034972 Arabidops
c 691	20	1.0	340000	9	AP001690	AP001690 Homo sapi	764	20	1.0	2586	8	BDICAR	BDICAR Arabidops
c 692	20	1.0	340000	9	AP001690	AP001690 Homo sapi	765	20	1.0	2606	3	BPE289770	BPE289770 Homo sapi
c 693	20	1.0	340000	9	AP001690	AP001690 Homo sapi	766	20	1.0	2687	8	AF024282	AF024282 Homo sapi
c 694	20	1.0	340000	9	AP001690	AP001690 Homo sapi	767	20	1.0	2787	9	PFATUBB	PFATUBB Homo sapi
c 695	20	1.0	340000	9	AP001690	AP001690 Homo sapi	768	20	1.0	2833	3	BC001427	BC001427 Homo sapi
c 696	20	1.0	340000	9	AP001690	AP001690 Homo sapi	769	20	1.0	2845	9	AY065425	AY065425 Arabidops
c 697	20	1.0	340000	9	AP001690	AP001690 Homo sapi	770	20	1.0	2851	8	BC021881	BC021881 Mus muscu
c 698	20	1.0	340000	9	AP001690	AP001690 Homo sapi	771	20	1.0	2887	10	ES3444	ES3444 SBEEN1 pol
c 699	20	1.0	340000	9	AP001690	AP001690 Homo sapi	772	20	1.0	2894	6	AY064013	AY064013 Arabidops
c 700	20	1.0	340000	9	AP001690	AP001690 Homo sapi	773	20	1.0	2908	10	AF159455	AF159455 Mus muscu
c 701	20	1.0	340000	9	AP001690	AP001690 Homo sapi	774	20	1.0	3340	8	ATPMKNT10A	ATPMKNT10A Homo sapi
c 702	20	1.0	340000	9	AP001690	AP001690 Homo sapi	775	20	1.0	3415	9	HSMB03569	HSMB03569 Homo sapi
c 703	20	1.0	340000	9	AP001690	AP001690 Homo sapi	776	20	1.0	3454	8	EL17067	EL17067 Human mRNa
c 704	20	1.0	340000	9	AP001690	AP001690 Homo sapi	777	20	1.0	3763	6	EL17068	EL17068 Human mRNa
c 705	20	1.0	340000	9	AP001690	AP001690 Homo sapi	778	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 706	20	1.0	340000	9	AP001690	AP001690 Homo sapi	779	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 707	20	1.0	340000	9	AP001690	AP001690 Homo sapi	780	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 708	20	1.0	340000	9	AP001690	AP001690 Homo sapi	781	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 709	20	1.0	340000	9	AP001690	AP001690 Homo sapi	782	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 710	20	1.0	340000	9	AP001690	AP001690 Homo sapi	783	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 711	20	1.0	340000	9	AP001690	AP001690 Homo sapi	784	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 712	20	1.0	340000	9	AP001690	AP001690 Homo sapi	785	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 713	20	1.0	340000	9	AP001690	AP001690 Homo sapi	786	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 714	20	1.0	340000	9	AP001690	AP001690 Homo sapi	787	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 715	20	1.0	340000	9	AP001690	AP001690 Homo sapi	788	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 716	20	1.0	340000	9	AP001690	AP001690 Homo sapi	789	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 717	20	1.0	340000	9	AP001690	AP001690 Homo sapi	790	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 718	20	1.0	340000	9	AP001690	AP001690 Homo sapi	791	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 719	20	1.0	340000	9	AP001690	AP001690 Homo sapi	792	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 720	20	1.0	340000	9	AP001690	AP001690 Homo sapi	793	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 721	20	1.0	340000	9	AP001690	AP001690 Homo sapi	794	20	1.0	3763	6	AB003062	AB003062 Homo sapi
c 722	20	1.0	340000	9	AP001690	AP001690 Homo sapi	795	20	1.0	3763	6	AB003062	AB003062 Homo sapi

796	20	1.0	6123	9	HSMB04590	AL833277	Homo sapi	c 865	870	20	1.0	74313	8	AP004977	AP004977	Lotus jap
797	20	1.0	7138	6	AX323841	AX323841	Sequence	c 870	870	20	1.0	75756	8	AP068601	AP068601	Homo sapi
C 798	20	1.0	11102	5	AF324888	AF324888	Homo sapi	c 871	871	20	1.0	77636	8	AB077832	AB077832	Arabidops
C 799	20	1.0	12029	3	AE001381	AE001381	Plasmodi	c 872	872	20	1.0	78578	2	PFMAL15P9	PFMAL15P9	Arabidops
C 800	20	1.0	12029	3	AE001412	AE001412	Plasmodi	c 873	873	20	1.0	78770	2	PFMAL15P9	PFMAL15P9	Arabidops
C 801	20	1.0	12029	3	AE001427	AE001427	Plasmodi	c 874	874	20	1.0	79457	2	AC034230	AC034230	Homo sapi
C 802	20	1.0	12229	3	AE001377	AE001377	Plasmodi	c 875	875	20	1.0	79769	2	AC098396	AC098396	Homo sapi
C 803	20	1.0	12355	2	AC116987	AC116987	Dictyoste	c 876	876	20	1.0	80030	2	AC115678	AC115678	Dictyoste
C 804	20	1.0	14202	2	AF324893S3	AF324893	Homo sapi	c 877	877	20	1.0	80845	2	AC097127	AC097127	Homo sapi
C 805	20	1.0	14613	3	AE001396	AE001396	Plasmodi	c 878	878	20	1.0	81177	9	AC016988	AC016988	Homo sapi
C 806	20	1.0	15342	3	AE001420	AE001420	Plasmodi	c 879	879	20	1.0	82411	8	AC068809	AC068809	Homo sapi
C 807	20	1.0	15387	1	SC5C11	SC5C11	Streptomy	c 880	880	20	1.0	82875	8	AC013482	AC013482	Arabidops
C 808	20	1.0	15903	2	AC115612	AC115612	Dictyoste	c 881	881	20	1.0	84162	8	T22J18	T22J18	Arabidops
C 809	20	1.0	15903	2	AC115612	AC115612	Dictyoste	c 882	882	20	1.0	84162	8	T22J18	T22J18	Arabidops
C 810	20	1.0	19369	2	AC133308	AC133308	Homo sapi	c 883	883	20	1.0	89703	9	AL356134	AL356134	Human DNA
C 811	20	1.0	21882	10	AL606782	AL606782	Mouse DNA	c 884	884	20	1.0	90289	2	AC107601	AC107601	Homo sapi
C 812	20	1.0	23324	2	AC020471	AC020471	Drosophi	c 885	885	20	1.0	91013	2	AC109943	AC109943	Rattus no
C 813	20	1.0	24053	9	ATY12227	ATY12227	Arabidops	c 886	886	20	1.0	92653	2	PFMAL1P1_3	PFMAL1P1_3	Continuation (4 of
C 814	20	1.0	24300	9	AF448221	AF448221	Homo sapi	c 887	887	20	1.0	92832	8	AC005917	AC005917	Arabidops
C 815	20	1.0	25117	2	AC116102	AC116102	Dictyoste	c 888	888	20	1.0	93329	2	AC097353	AC097353	Rattus no
C 816	20	1.0	25339	3	AB017765	AB017765	Trypanoso	c 889	889	20	1.0	93491	2	AC116967	AC116967	Dictyoste
C 817	20	1.0	25499	2	AC017653	AC017653	Drosophi	c 890	890	20	1.0	93681	9	AC107463	AC107463	Homo sapi
C 818	20	1.0	27744	2	AC019881	AC019881	Drosophi	c 891	891	20	1.0	93805	5	HS424E5	HS424E5	Human DNA
C 819	20	1.0	27934	9	AC106717	AC106717	Homo sapi	c 892	892	20	1.0	94292	9	AC108107	AC108107	Homo sapi
C 820	20	1.0	29001	2	AC018796	AC018796	Drosophi	c 893	893	20	1.0	95175	2	AC117861	AC117861	Rattus no
C 821	20	1.0	30220	2	AC019816	AC019816	Drosophi	c 894	894	20	1.0	95996	8	ATP18P9	ATP18P9	Arabidops
C 822	20	1.0	31520	9	HSX1ST1	HSX1ST1	Human X413	c 895	895	20	1.0	97034	8	AC005698	AC005698	Genomic s
C 823	20	1.0	31812	8	DMBNA3K23	DMBNA3K23	S.pombe c	c 896	896	20	1.0	99680	2	AC009769	AC009769	Homo sapi
C 824	20	1.0	32391	3	AC003690	AC003690	Human Chr	c 897	897	20	1.0	99964	2	RN364I06	RN364I06	Homo sapi
C 825	20	1.0	34052	9	AC129092	AC129092	Medicago	c 898	898	20	1.0	100000	9	AB020859	AB020859	Rattus no
C 826	20	1.0	34096	2	AC116958	AC116958	Dictyoste	c 899	899	20	1.0	100000	9	AB020854	AB020854	Homo sapi
C 827	20	1.0	35119	9	HS27C3	HS27C3	Human DNA	c 900	900	20	1.0	100000	9	AB020872	AB020872	Homo sapi
C 828	20	1.0	35700	9	AC013004	AC013004	Drosophi	c 901	901	20	1.0	100000	9	AP000210	AP000210	Homo sapi
C 829	20	1.0	36576	2	AC011529	AC011529	Homo sapi	c 902	902	20	1.0	100000	17	AP000132	AP000132	Homo sapi
C 830	20	1.0	37604	9	AC011548	AC011548	Homo sapi	c 903	903	20	1.0	100282	8	AC006298	AC006298	Rattus no
C 831	20	1.0	38351	6	AX244215	AX244215	Sequence	c 904	904	20	1.0	101213	8	AC113851	AC113851	Rattus no
C 832	20	1.0	38653	9	AC012633	AC012633	Homo sapi	c 905	905	20	1.0	101644	8	AC006287	AC006287	Arabidops
C 833	20	1.0	40965	9	AC090631	AC090631	Homo sapi	c 906	906	20	1.0	102326	2	AC120734	AC120734	Rattus no
C 834	20	1.0	41135	9	AF217651	AF217651	Drosophi	c 907	907	20	1.0	102422	2	AC129557	AC129557	Mus muscu
C 835	20	1.0	45142	3	RN363K5	RN363K5	Rattus no	c 908	908	20	1.0	103922	2	AC096546	AC096546	Homo sapi
C 836	20	1.0	47101	2	AC115593	AC115593	Dictyoste	c 909	909	20	1.0	103942	2	AC091859	AC091859	Homo sapi
C 837	20	1.0	47791	2	AC116305	AC116305	Dictyoste	c 910	910	20	1.0	103989	9	HSDA109P7	HSDA109P7	Human DNA
C 838	20	1.0	47791	2	AC116305	AC116305	Dictyoste	c 911	911	20	1.0	104320	2	AC126914	AC126914	Rattus no
C 839	20	1.0	48352	9	AC0101079	AC0101079	Mus muscu	c 912	912	20	1.0	104351	9	AC0022419	AC0022419	Homo sapi
C 840	20	1.0	48871	8	AC082643	AC082643	Arabidops	c 913	913	20	1.0	104786	2	AC111859	AC111859	Rattus no
C 841	20	1.0	48891	8	AC103694	AC103694	Homo sapi	c 914	914	20	1.0	104786	2	AC005504	AC005504	Plasmodi
C 842	20	1.0	50711	2	AC116551	AC116551	Dictyoste	c 915	915	20	1.0	105036	2	AC073181	AC073181	Homo sapi
C 843	20	1.0	52514	2	AC122764	AC122764	Mus muscu	c 916	916	20	1.0	105306	8	ATP9D24	ATP9D24	Arabidops
C 844	20	1.0	54472	2	AC116963	AC116963	Dictyoste	c 917	917	20	1.0	105797	9	AC011345	AC011345	Homo sapi
C 845	20	1.0	56152	2	AC100895	AC100895	Mus muscu	c 918	918	20	1.0	106039	9	AL589786	AL589786	Human DNA
C 846	20	1.0	57455	2	AC115680	AC115680	Dictyoste	c 919	919	20	1.0	106144	2	AC116425	AC116425	Dictyoste
C 847	20	1.0	59004	9	AC079882	AC079882	Homo sapi	c 920	920	20	1.0	106144	2	AC116425	AC116425	Dictyoste
C 848	20	1.0	59427	9	D78344	D78344	Mouse DNA	c 921	921	20	1.0	106146	3	AC117080	AC117080	Dictyoste
C 849	20	1.0	59641	10	AC021776	AC021776	Homo sapi	c 922	922	20	1.0	106638	2	AC117983	AC117983	Rattus no
C 850	20	1.0	61643	2	AC015206	AC015206	Homo sapi	c 923	923	20	1.0	106638	2	AC117983	AC117983	Rattus no
C 851	20	1.0	62001	2	AC103865	AC103865	Homo sapi	c 924	924	20	1.0	106963	2	AC127409	AC127409	Rattus no
C 852	20	1.0	62497	2	AC104349	AC104349	Homo sapi	c 925	925	20	1.0	107658	9	AC119592	AC119592	Rattus no
C 853	20	1.0	64360	9	AC118039	AC118039	Mus muscu	c 926	926	20	1.0	108346	9	AC034067	AC034067	Homo sapi
C 854	20	1.0	64480	9	AC128496	AC128496	Homo sapi	c 927	927	20	1.0	108729	2	AC084298	AC084298	Rattus no
C 855	20	1.0	64719	2	AC109230	AC109230	Mus muscu	c 928	928	20	1.0	109290	2	HS838L14	HS838L14	Homo sapi
C 856	20	1.0	65353	2	AC099807	AC099807	Homo sapi	c 929	929	20	1.0	110000	2	AC112264_2	AC112264_2	Continuation (3 of
C 857	20	1.0	65493	2	AC102158	AC102158	Mus muscu	c 930	930	20	1.0	110000	2	AC112266_0	AC112266_0	Continuation (2 of
C 858	20	1.0	65710	2	AC025892	AC025892	Homo sapi	c 931	931	20	1.0	110000	2	PFMAL1P1_0	PFMAL1P1_0	Continuation (2 of
C 859	20	1.0	65711	3	PFMAL1P4	PFMAL1P4	Mus muscu	c 932	932	20	1.0	110000	2	PFMAL1P1_0	PFMAL1P1_0	Continuation (2 of
C 860	20	1.0	66441	3	AC108426	AC108426	Mus muscu	c 933	933	20	1.0	110000	2	PFMAL1P1_0	PFMAL1P1_0	Continuation (2 of
C 861	20	1.0	67358	2	AC079979	AC079979	Homo sapi	c 934	934	20	1.0	110000	2	PFMAL1P1_0	PFMAL1P1_0	Continuation (2 of
C 862	20	1.0	67540	2	AC101227	AC101227	Mus muscu	c 935	935	20	1.0	111298	2	RN39A23	RN39A23	Human DNA
C 863	20	1.0	68529	2	AC123638	AC123638	Mus muscu	c 936	936	20	1.0	112616	9	HS359615	HS359615	Homo sapi
C 864	20	1.0	70601	2	AC017306	AC017306	Drosophi	c 937	937	20	1.0	114630	2	AC025461	AC025461	Homo sapi
C 865	20	1.0	70627	2	AL663116	AL663116	Mouse DNA	c 938	938	20	1.0	114630	2	AC079628	AC079628	Homo sapi
C 866	20	1.0	70735	10	AP000575	AP000575	Homo sapi	c 939	939	20	1.0	114897	2	AP003624	AP003624	Oryza sat
C 867	20	1.0	73431	2	AL589762	AL589762	Human DNA	c 940	940	20	1.0	115489	2	AC117072	AC117072	Dictyoste
C 868	20	1.0	74089	9	AL589762	AL589762	Human DNA	c 941	941	20	1.0	116636	3	PFMAL1P3	PFMAL1P3	Plasmodi

C 942	20	1.0	116596	3	PFMAL3P3	298547	Plasmodium
C 943	20	1.0	116747	2	AC123329	AC10644	Rattus no
C 944	20	1.0	116897	2	AC104644	AC10940	Rattus no
C 945	20	1.0	117074	2	AC110940	AC10608	Rattus no
C 946	20	1.0	117585	2	AC108608	AC1353774	Human DNA
C 947	20	1.0	118588	9	AL353774	AC126302	Rattus no
C 948	20	1.0	119495	2	AC126302	CEY43P8C	Rattus no
C 949	20	1.0	130492	3	CEY43P8C	AL032637	Caenorhab
C 950	20	1.0	130592	2	AL714031	AL714031	Danio rer
C 951	20	1.0	130722	2	AC111994	AC111994	Rattus no
C 952	20	1.0	130857	9	AC116334	AC1068715	Homo sapi
C 953	20	1.0	131737	9	AC008715	AC008715	Homo sapi
C 954	20	1.0	132073	2	AC010112	AC010112	Drosophila
C 955	20	1.0	132252	2	AC109880	AC109880	Rattus no
C 956	20	1.0	132279	2	AL359845	AL359845	Human DNA
C 957	20	1.0	132279	2	AC119112	AC119112	Rattus no
C 958	20	1.0	132637	9	AC006062	AC006062	Homo sapi
C 959	20	1.0	133260	10	AC087041	AC087041	Rattus no
C 960	20	1.0	133280	2	AC117076	AC117076	Dicystoste
C 961	20	1.0	141114	8	HSJ858M42	AL118510	Human DNA
C 962	20	1.0	142320	10	AC005259	AC005259	Mouse BAC
C 963	20	1.0	142321	8	OSA245900	AJ245900	Oryza sat
C 964	20	1.0	150666	2	AC074286	AC074286	Homo sapi
C 965	20	1.0	157522	2	AC015695	AC015695	Homo sapi
C 966	20	1.0	157168	9	HS116301	AL031589	Human DNA
C 967	20	1.0	157423	2	AC117018	AC117018	Rattus no
C 968	20	1.0	130012	2	AC120632	AC120632	Rattus no
C 969	20	1.0	130369	9	AC074090	AC074090	Homo sapi
C 970	20	1.0	130403	9	AL158139	AL158139	Human DNA
C 971	20	1.0	131338	9	AC099798	AC099798	Homo sapi
C 972	20	1.0	131353	9	HS508115	AL021707	Human DNA
C 973	20	1.0	131736	2	AC021527	AC021527	Homo sapi
C 974	20	1.0	131919	9	AC008700	AC008700	Homo sapi
C 975	20	1.0	132000	2	AC116976	AC116976	Dicystoste
C 976	20	1.0	132035	2	AC125548	AC125548	Rattus no
C 977	20	1.0	132655	3	AC010015	AC010015	Drosophila
C 978	20	1.0	133275	2	AC107956	AC107956	Homo sapi
C 979	20	1.0	133279	9	AC012317	AC012317	Homo sapi
C 980	20	1.0	133348	9	AC114322	AC114322	Homo sapi
C 981	20	1.0	133445	2	AC073440	AL022233	Human DNA
C 982	20	1.0	134137	9	HS3E5	AL022233	Human DNA
C 983	20	1.0	134319	2	AC068673	AC068673	Homo sapi
C 984	20	1.0	135123	9	AC079950	AC079950	Homo sapi
C 985	20	1.0	135439	2	AC080065	AC080065	Homo sapi
C 986	20	1.0	137060	2	AC116867	AC116867	Mus muscu
C 987	20	1.0	137077	2	AC007562	AC007562	Homo sapi
C 988	20	1.0	137942	10	AL645745	AL645745	Mouse DNA
C 989	20	1.0	138151	2	AC094070	AC094070	Rattus no
C 990	20	1.0	138264	9	AC008722	AC008722	Homo sapi
C 991	20	1.0	139736	2	AC112779	AC112779	Rattus no
C 992	20	1.0	140133	2	AC091050	AC091050	Homo sapi
C 993	20	1.0	140583	9	AC010978	AC010978	Homo sapi
C 994	20	1.0	140702	2	AC068420	AC068420	Homo sapi
C 995	20	1.0	140952	2	AC079982	AC079982	Homo sapi
C 996	20	1.0	141790	9	AL590636	AL590636	Human DNA
C 997	20	1.0	141899	9	AC006334	AC006334	Homo sapi
C 998	20	1.0	143331	9	AC091214	AC091214	Homo sapi
C 999	20	1.0	144515	9	AC005831	AC005831	Homo sapi
C1000	20	1.0	144882	2	AC008726	AC008726	Homo sapi

## ALIGNMENTS

RESULT 1							
AR203335	AR203335	2032 bp	DNA	linear	PAT	20-JUN-2002	
LOCUS	Sequence 2 from patent US 656365.						
DEFINITION	AR203335						
ACCESSION	AR203335.1	GI:21499698					
VERSION							
KEYWORDS							
SOURCE	Unknown.						
ORGANISM	Unknown.						

REFERENCE	1 (bases 1 to 2032)	Unclassified.
AUTHORS	Bistrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S.	
TITLE	Method of determining whether an agent modulates glycosyl	
JOURNAL	sulfotransferase-3	
FEATURES	Patent: US 636365-A 2 02-APR-2002;	
source	Location/Qualifiers	
BASE COUNT	468 a 569 c 490 g 505 t	
ORIGIN		
Query Match	100.0%; Score 2032; DB 6; Length 2032;	
Best local similarity	100.0%; Pred. No. 0;	
Matches 2032; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
1	GGCTCAGGCGCCAGATGCTCTCCAGTGTGGGGGAAATGCTTCTATTGCTTCCAG	60
1	GGCTCAGGCGCCAGATGCTCTCCAGTGTGGGGGAAATGCTTCTATTGCTTCCAG	60
61	CCACCTCAGAGAGTCTCCACCCCTGAGTCTCAGAGTGTAAAGTGTATTTCA	120
61	CCACCTCAGAGAGTCTCCACCCCTGAGTCTCAGAGTGTAAAGTGTATTTCA	120
61	CCACCTCAGAGAGTCTCCACCCCTGAGTCTCAGAGTGTAAAGTGTATTTCA	120
121	CAGCTTCTGAGGAGGAGTCTTCTCAAGCCGCTTGTCAAGGCTTCCACTGACAC	180
121	CAGCTTCTGAGGAGGAGTCTTCTCAAGCCGCTTGTCAAGGCTTCCACTGACAC	180
121	CAGCTTCTGAGGAGGAGTCTTCTCAAGCCGCTTGTCAAGGCTTCCACTGACAC	180
181	AATGCTACGCTTCAAAAATGATGCTCTGCTTCTGCTTCCAGATGGCATTT	240
181	AATGCTACGCTTCAAAAATGATGCTCTGCTTCTGCTTCCAGATGGCATTT	240
241	GGCTATCTTCTCCACATGATGACGACACATGATGCTCTGCTTATGAAGGACAC	300
241	GGCTATCTTCTCCACATGATGACGACACATGATGCTCTGCTTATGAAGGACAC	300
241	GGCTATCTTCTCCACATGATGACGACACATGATGCTCTGCTTATGAAGGACAC	300
301	CGAGGCGATGACGCTGCTGCTTCTTCTCCGCGCTTCTTCTTTGAGGACA	360
301	CGAGGCGATGACGCTGCTGCTTCTTCTCCGCGCTTCTTCTTTGAGGACA	360
301	CGAGGCGATGACGCTGCTGCTTCTTCTCCGCGCTTCTTCTTTGAGGACA	360
361	GGCTTTTGGGAGACACCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	420
361	GGCTTTTGGGAGACACCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	420
361	GGCTTTTGGGAGACACCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	420
421	GACCTTCAAGGAGACACCCGCTGATGCTGACATGCTGCTGCTGCTGCTGCTG	480
421	GACCTTCAAGGAGACACCCGCTGATGCTGACATGCTGCTGCTGCTGCTGCTG	480
421	GACCTTCAAGGAGACACCCGCTGATGCTGACATGCTGCTGCTGCTGCTGCTG	480
481	GGCTTTTGGGAGACACCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	540
481	GGCTTTTGGGAGACACCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	540
481	GGCTTTTGGGAGACACCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	540
541	GTCAGACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	600
541	GTCAGACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	600
541	GTCAGACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	600
601	GATCCCAAGATGAATATATATATATATATATATATATATATATATATATAT	660
601	GATCCCAAGATGAATATATATATATATATATATATATATATATATATATAT	660
601	GATCCCAAGATGAATATATATATATATATATATATATATATATATATATAT	660
661	CTTTGAGTGTGTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	720
661	CTTTGAGTGTGTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	720
661	CTTTGAGTGTGTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	720
721	CTTTTCAACCTGCACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	780
721	CTTTTCAACCTGCACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	780
721	CTTTTCAACCTGCACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	780
781	CGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	840
781	CGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	840
781	CGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	840
841	TCTCATGATGACAGTGCATTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAG	900

Db	841	TCTCATGATGACAGTGGCAATTGTGATGCGGCCAGCATGAGCAGAAACTCAAGAGAGATA	900
QY	901	CCAAOCCCTACTATGTGATGCAAGCTCATCTGCCAAAGCCAGCTGGAGATCTACAGACAT	960
Db	901	CCAAOCCCTACTATGTGATGCAAGCTCATCTGCCAAAGCCAGCTGGAGATCTACAGACAT	960
QY	961	CCAGGCTCTGGCCCAAGGCCCTGCGAGGAAGCTCACTGCTGGTGGCTATGAGAGCTGGC	1020
Db	961	CCAGGCTCTGGCCCAAGGCCCTGCGAGGAAGCTCACTGCTGGTGGCTATGAGAGAGCTGGC	1020
QY	1021	TCGAGCCCCCTGTGGCCCCAGACTTCCGCAATGTATGAAATTCGTGGGATTTGGAATTCCTGGC	1080
Db	1021	TCGAGCCCCCTGTGGCCCCAGACTTCCGCAATGTATGAAATTCGTGGGATTTGGAATTCCTGGC	1080
QY	1081	CCATCTTCAGACCTGGGGTGGCATACATCACCAGGCGAGGGGATGAGGTGGACACGCTTT	1140
Db	1081	CCATCTTCAGACCTGGGGTGGCATACATCACCAGGCGAGGGGATGAGGTGGACACGCTTT	1140
QY	1141	CCACACAATATCCAGGGGATCCCTTATATGTCTCCAGGCTTGGCGCTGGTCTTTGCCCTA	1200
Db	1141	CCACACAATATCCAGGGGATCCCTTATATGTCTCCAGGCTTGGCGCTGGTCTTTGCCCTA	1200
QY	1201	TGAAAAGSTTTCTCGACTTCAGAAAGCTTGTGGCGATGCCATGATTAATTTCTGGGCTACCG	1260
Db	1201	TGAAAAGSTTTCTCGACTTCAGAAAGCTTGTGGCGATGCCATGATTAATTTCTGGGCTACCG	1260
QY	1261	CCACGTCGATCTGACACAGACAGAGAAACCTGTGGGATCTGTCTGTACCTGACGGAC	1320
Db	1261	CCACGTCGATCTGACACAGACAGAGAAACCTGTGGGATCTGTCTGTACCTGACGGAC	1320
QY	1321	TGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTTCGTCGCCACCTGGTGTGAGCC	1380
Db	1321	TGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTTCGTCGCCACCTGGTGTGAGCC	1380
QY	1381	TCAGTCACCTTCTGTGAAGCTTCTGAGCCTTCCCTACACTCTGTAGCCTTAACATCATG	1440
Db	1381	TCAGTCACCTTCTGTGAAGCTTCTGAGCCTTCCCTACACTCTGTAGCCTTAACATCATG	1440
QY	1441	TCTGTGGTATCACACAGAGTGTGTGATTTGTCCACAGTGTCAAGCAGAAAGGACTTTT	1500
Db	1441	TCTGTGGTATCACACAGAGTGTGTGATTTGTCCACAGTGTCAAGCAGAAAGGACTTTT	1500
QY	1501	GTGTCCATGCTTGTGTCTAGAAAAACAGCTGGGGAACTTATGTAGCGACACATCCAC	1560
Db	1501	GTGTCCATGCTTGTGTCTAGAAAAACAGCTGGGGAACTTATGTAGCGACACATCCAC	1560
QY	1561	CAGTGAACACAGGATATGTGCTTCTCTTTTCTTGAATCTTCGCTGGGACAGACTTGG	1620
Db	1561	CAGTGAACACAGGATATGTGCTTCTCTTTTCTTGAATCTTCGCTGGGACAGACTTGG	1620
QY	1621	AGACTTTTGTGCTGCGAGGCCATTAAAGCAGACAGATACAGTATCGTAGTTATCATATA	1680
Db	1621	AGACTTTTGTGCTGCGAGGCCATTAAAGCAGACAGATACAGTATCGTAGTTATCATATA	1680
QY	1681	ACCTCCCGTGCACATCTTGGCCCAATGGGGAATGGATCTTTACCCAAAGAGCTCCAGC	1740
Db	1681	ACCTCCCGTGCACATCTTGGCCCAATGGGGAATGGATCTTTACCCAAAGAGCTCCAGC	1740
QY	1741	ATTTTCCACAGAGATGCANAATTCGAGCCCTTGGAGTTCGCAGTGAATTCAGAGAACTA	1800
Db	1741	ATTTTCCACAGAGATGCANAATTCGAGCCCTTGGAGTTCGCAGTGAATTCAGAGAACTA	1800
QY	1801	GTGGGAACAAAGTGGATGCCCTATTATGAGCTTGAACATCAAGAGTATCCGTAATCAGA	1860
Db	1801	GTGGGAACAAAGTGGATGCCCTATTATGAGCTTGAACATCAAGAGTATCCGTAATCAGA	1860
QY	1861	AATATGAACAAATCTCTGACAAAAGACAGACGCTTAAAGTTCACAGGCTGGCTGGCC	1920
Db	1861	AATATGAACAAATCTCTGACAAAAGACAGACGCTTAAAGTTCACAGGCTGGCTGGCC	1920
QY	1921	TGCATTTGAATATACCTTCCCTGTGCAATTTCCCATCAGATGAAGACTTTGACCTGTG	1980

DB	1921	TGCATTGTGAATTCACCTTCCCTCTGCATTTTCCATCATAGAAAGACTTTGACCTGTG	1980
QY	1981	AAGCTGCCACTCTGTTAATPCTAAATTCGCCAATTAGAAAAAATTTTAAAAA	2032
DB	1981	AAGCTGCCACTCTGTTAATPCTAAATTCGCCAATTAGAAAAAATTTTAAAAA	2032
RESULT 2			
AF131235			
LOCUS	AF131235	2032 bp	mRNA
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,		
ACCESSION	AF131235		
VERSION	AF131235.1		
KEYWORDS	GI:4927113		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2032)		
TITLE	Bistrup, A., Bhakta, S., Lee, J.-K., Below, Y.Y., Gunn, M.D., Zuo, F.R., Huang, C.C., Kannagi, R., Rosen, S.D. and Hemmerich, S.		
JOURNAL	Reconstituted expression of two specificities function in the sulfotransferase of high endothelial cell ligands for L-selectin		
PUBMED	J. Cell Biol. 145 (4), 899-910 (1999)		
REFERENCE	9264336		
AUTHORS	10330415		
TITLE	2 (bases 1 to 2032)		
JOURNAL	Bistrup, A., Tangemann, K., Bhakta, S., Lee, J.-K., Below, Y.Y., Gunn, M.D., Zuo, F.-R., Huang, C.-C., Kannagi, R., Rosen, S.D. and Hemmerich, S.		
FEATURES	Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA		
Source	Location/Qualifiers		
	1..2032		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_type="high endothelial"		
	/tissue_type="tonsil"		
	182..1342		
	/function="biosynthesis of endothelial ligands for L-selectin"		
	/note="HEC-GlcNAc-6-ST"		
	/cdoon_start=1		
	/product="N-acetylglucosamine 6-O-sulfotransferase"		
	/protein_id="AAD3015.1"		
	/db_xref="GI:4927114"		
	/translation="MLPKKKLLFLVSKAATLALPFRHNSITLSMKAGPERKHHVLISSMRSSSEYGLRQHPDVFYLERPMHWMTFQSTAMLHVAVDLIRAVFLCDVSEFDAYMEPGPRKSSLFQWENSLCASAQCILIPQDELIPAHCRLLSQDPLEVEKACRSYSHVVLKEKRSFLNQSLEPLKDPSSLNIHYLWDRPRAVRSERKGDMLIDSRITVMGQEHOKLKEDDOPYVMQVICOSELEYKTIQSLPKALOERYLLVREEDLRAPIAYQTSRMVEYFGLFPHLOTWVHNITRGKMGDGHATHTAADAINTVSQAMRMSLPYEVSRSLQKACGDAMNLLGYRHRSRSDQRNLLDLSTWYPEQIH"		
BASE COUNT	466 a 569 c 490 g 505 t		
ORIGIN			
Query Match	100.0%; Score 2032; DB 9; Length 2032;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 2032;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	GGCTTGAGGCGGAGATGCTCCAGTCTGGGGGAAAAATGCTTCATTTGCTTCCAG	60
DB	1	GGCTTGAGGCGGAGATGCTCCAGTCTGGGGGAAAAATGCTTCATTTGCTTCCAG	60
QY	61	CCACCTCAGAGAGTCTCCCAACCCCTTAGTCTAGCAGTGTAAAGCTGTACTTCA	120
DB	61	CCACCTCAGAGAGTCTCCCAACCCCTTAGTCTAGCAGTGTAAAGCTGTACTTCA	120
QY	121	CAGCTTCTGGAGGAGAGTCTTCTCAAGCCGCTTTCAGAGGTCTTCACCTAGCAG	180
DB	121	CAGCTTCTGGAGGAGAGTCTTCTCAAGCCGCTTTCAGAGGTCTTCACCTAGCAG	180

QY 181 AATGCTACTGGCTAAAAAATGAAGCTGCTGTTTCTGTTTCCAGATGGCATCTT 240  
DB 181 AATGCTACTGGCTAAAAAATGAAGCTGCTGTTTCTGTTTCCAGATGGCATCTT 240  
QY 241 GGCTCTATTTCTCCATGATGACAGCCACACATCAGCTCCCTGTCTATGAAGGCACAGCC 300  
DB 241 GGCTCTATTTCTCCATGATGACAGCCACACATCAGCTCCCTGTCTATGAAGGCACAGCC 300  
QY 301 CGAGGGCATGACGCTGGTGTCTGTTCTGCTGGGCGCTGGCTGCTTCTTTGGGGGCA 360  
DB 301 CGAGGGCATGACGCTGGTGTCTGTTCTGCTGGGCGCTGGCTGCTTCTTTGGGGGCA 360  
QY 361 GCTTTTGGGAGACCCAGATGTTTCTTACCTGATGAGGCCCGCTGGGACGTGTGGAT 420  
DB 361 GCTTTTGGGAGACCCAGATGTTTCTTACCTGATGAGGCCCGCTGGGACGTGTGGAT 420  
QY 421 GACCTTCAAGAGACAGCCGCTGGATGCTGACATGGCTGTGGGATGATGACGGC 480  
DB 421 GACCTTCAAGAGAGACAGCCGCTGGATGCTGACATGGCTGTGGGATGATGACGGC 480  
QY 481 GCTCTCTTGTGACACATGAGCGTCTTGTGATGCTACATGGAACCTGGTCCCGGAGACA 540  
DB 481 GCTCTCTTGTGACACATGAGCGTCTTGTGATGCTACATGGAACCTGGTCCCGGAGACA 540  
QY 541 GTCCAGCTCTTTCAAGTGGGAGAACAGCCGGCCCTGTGTTCTGCACTGCTGTGACAT 600  
DB 541 GTCCAGCTCTTTCAAGTGGGAGAACAGCCGGCCCTGTGTTCTGCACTGCTGTGACAT 600  
QY 601 CATCCCAAGATGAATCATCCCGGGGCTCACTGAGGCTCTGTGAGTCAAGAGCC 660  
DB 601 CATCCCAAGATGAATCATCCCGGGGCTCACTGAGGCTCTGTGAGTCAAGAGCC 660  
QY 661 CTTTGAGTGTGGAGAGGCTCCGCTCTCAAGCCAGCTGTGCTCAAGAGAGTGGC 720  
DB 661 CTTTGAGTGTGGAGAGGCTCCGCTCTCAAGCCAGCTGTGCTCAAGAGAGTGGC 720  
QY 721 CTTCTCAACCGAGAGTCCCTCAACCGCTGTGAAGAACCCCTCCCAACCGCATAT 780  
DB 721 CTTCTCAACCGAGAGTCCCTCAACCGCTGTGAAGAACCCCTCCCAACCGCATAT 780  
QY 781 CCGGACCTGGTCCGGGAGCCCGGGCGGTTCGTTCCGAGAGAGCAAAAGGAGA 840  
DB 781 CCGGACCTGGTCCGGGAGCCCGGGCGGTTCGTTCCGAGAGAGCAAAAGGAGA 840  
QY 841 TCTCATGATGACAGTGGCATTTGATGGGGGACATGAGAGAACTCAAGAGAGGA 900  
DB 841 TCTCATGATGACAGTGGCATTTGATGGGGGACATGAGAGAACTCAAGAGAGGA 900  
QY 901 CCAACCCCTACTATGTGATGACAGTCACTGCCAAGCCAGCTGAGATCTACAAGACAT 960  
DB 901 CCAACCCCTACTATGTGATGACAGTCACTGCCAAGCCAGCTGAGATCTACAAGACAT 960  
QY 961 CCAAGTCTTGGCCCAAGGCGCTGCAAGAAAGCTACCTGCTGTGGCTATGAGAGCTGGC 1020  
DB 961 CCAAGTCTTGGCCCAAGGCGCTGCAAGAAAGCTACCTGCTGTGGCTATGAGAGCTGGC 1020  
QY 1021 TCGAGCCCTTGGGCGCCAGACTTCCCGAATGATGAATGATGGATTTGGAATCTTGGC 1080  
DB 1021 TCGAGCCCTTGGGCGCCAGACTTCCCGAATGATGAATGATGGATTTGGAATCTTGGC 1080  
QY 1081 CCACTCTCAGACCTGGGTGATTAACATCACCCGAGGCAAGGCGATGGGTACCAAGCTTT 1140  
DB 1081 CCACTCTCAGACCTGGGTGATTAACATCACCCGAGGCAAGGCGATGGGTACCAAGCTTT 1140  
QY 1141 CCACACAAATGCCAGGATGCCCTTATGCTCCAGAGGCTGGGCTGTGGCTTTGGCCCTA 1200  
DB 1141 CCACACAAATGCCAGGATGCCCTTATGCTCCAGAGGCTGGGCTGTGGCTTTGGCCCTA 1200  
QY 1201 TGAAGAAGTTTCTGACATTCAGAAAGCTGTGGAGATGACATGAATTTGCTGGGCTACGC 1260  
DB 1201 TGAAGAAGTTTCTGACATTCAGAAAGCTGTGGAGATGACATGAATTTGCTGGGCTACGC 1260

QY 1261 CCACGTCAAGTCTGAACAGAGAGAAAGCAAGCAAGCTGTTGCTGGAATCTCTGTAACCTGGAC 1320  
DB 1261 CCACGTCAAGTCTGAACAGAGAGAGAAAGCAAGCAAGCTGTTGCTGGAATCTCTGTAACCTGGAC 1320  
QY 1321 TGTCCCTGAGCAATTCACCTAGAGGGTTGAGAGGCTTGTGCTGCCACCTGGTGTACGC 1380  
DB 1321 TGTCCCTGAGCAATTCACCTAGAGGGTTGAGAGGCTTGTGCTGCCACCTGGTGTACGC 1380  
QY 1381 TCACTACATTTCTGAAATGCTTCTGAGCTTGGCCTACATCTCTGAGCTTAACATACAT 1440  
DB 1381 TCACTACATTTCTGAAATGCTTCTGAGCTTGGCCTACATCTCTGAGCTTAACATACAT 1440  
QY 1441 TCTGTGGATCACTAGATGTGATGTGTCCACAGCTGCTCAAGCAGAGAGACTTTT 1500  
DB 1441 TCTGTGGATCACTAGATGTGATGTGTGTCCACAGCTGCTCAAGCAGAGAGACTTTT 1500  
QY 1501 GTGTCCATGCTGTGTGTGAAGAACAGATGGGAGACTTAATGTAGAGCAGCATCTCCAC 1560  
DB 1501 GTGTCCATGCTGTGTGTGAAGAACAGATGGGAGACTTAATGTAGAGCAGCATCTCCAC 1560  
QY 1561 CAGTGAACAGGGTATGCTCTTCTTCTTCTTCTGATCTTCTGCTGCTGGCAGACTTCAG 1620  
DB 1561 CAGTGAACAGGGTATGCTCTTCTTCTTCTTCTGATCTTCTGCTGCTGGCAGACTTCAG 1620  
QY 1621 AGACTTTGTGGCTTGAGAGGCTATTAAGCAGACAGATATAGTGAATATGCTATTA 1680  
DB 1621 AGACTTTGTGGCTTGAGAGGCTATTAAGCAGACAGATATAGTGAATATGCTATTA 1680  
QY 1681 ACCTCCCTGCCACATCTTGCCCAATGAGGGAATGATCTTTCACCAAGAGCTCACAGC 1740  
DB 1681 ACCTCCCTGCCACATCTTGCCCAATGAGGGAATGATCTTTCACCAAGAGCTCACAGC 1740  
QY 1741 ATTTTCCACAGATGCAAAATTTGAGCCCTTGTGAGCTTCCAGTGGATTTCAAGAGAGA 1800  
DB 1741 ATTTTCCACAGATGCAAAATTTGAGCCCTTGTGAGCTTCCAGTGGATTTCAAGAGAGA 1800  
QY 1801 GTGGGAGCAAGGTGGATGCTACTATGAGCTGACATCAAGATATGCTATTAATTCAGA 1860  
DB 1801 GTGGGAGCAAGGTGGATGCTACTATGAGCTGACATCAAGATATGCTATTAATTCAGA 1860  
QY 1861 AATATGAACAAATCTGACACAAAGAGCAAGCTTAACTTCAACAGAGCTGCTGGGC 1920  
DB 1861 AATATGAACAAATCTGACACAAAGAGCAAGCTTAACTTCAACAGAGCTGCTGGGC 1920  
QY 1921 TGCATTGAATATCACTTCCCTGCTGATTTCCATGCAATGAGAAAGATTTGACCTGTG 1980  
DB 1921 TGCATTGAATATCACTTCCCTGCTGATTTCCATGCAATGAGAAAGATTTGACCTGTG 1980  
QY 1981 AAGCTGCATCTGTTAATCTAAATTTCCCAATTAAGAAAAA 2032  
DB 1981 AAGCTGCATCTGTTAATCTAAATTTCCCAATTAAGAAAAA 2032

RESULT 3  
AK026635  
LOCUS AK026635 2011 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens CDNA: FLJ22982 fls, clone KAT11454, highly similar to AF11235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.  
ACCESSION AK026635  
VERSION AK026635.1 GI:10439531  
KEYWORDS oligo capping: fls (full insert sequence).  
SOURCE Homo sapiens signet-ring cell carcinoma cell line:KATO III CDNA to mRNA, clone.lib:KAT clone:KAT11454.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NED human cdna sequencing project  
JOURNAL Unpublished



REFERENCE	2 (bases 1 to 111)	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
AUTHORS	Submitted (29-AUG-2000)	Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp; Tel:81-3-5449-5286, Fax:81-3-5449-5416)
TITLE	NEBO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
JOURNAL		
COMMENT		
FEATURES	Location/Qualifiers	
Source	1..2011	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="KAT11454"
		/cell_line="KATO III"
		/cell_type="signet-ring cell carcinoma"
		/clone_lib="KAT"
		/note="cloning vector pME18SFL3"
Misc_feature	1..2011	/note="highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT	482 a	541 c 475 g 513 t
ORIGIN		
Query Match	91.3%; Score 1856; DB 9; Length 2011;	
Best Local Similarity	99.9%; Pred. No. 0;	
Matches 1906; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	111 GTTACTTTACAGCTTCTCTGGAGACGAGTCTTTTCAGCCGCTCTGCAAGCTCTCC	170
Db	1 GTTACTTTACAGCTTCTCTGGAGACGAGTCTTTTCAGCCGCTCTGCAAGCTCTCC	60
QY	171 ACTTCAGACAATGCTACTGCTCTAAATAATGAAGCTCCTGTTCTGTTTCCAGA	230
Db	61 ACTTCAGACAATGCTACTGCTCTAAATAATGAAGCTCCTGTTTCTGTTTCCAGA	120
QY	231 TGGCATTCTGGCTCTATCTTCCATGTACAGCCACAACATCAGCTCCCTGTATGA	290
Db	121 TGGCATTCTGGCTCTATCTTCCATGTACAGCCACAACATCAGCTCCCTGTATGA	180
QY	291 AGGCACAGCCGAGCGCATGAGTGTGTTCTTCTGCGGCTGTGGCTCTTCTT	350
Db	181 AGGCACAGCCGAGCGCATGAGTGTGTTCTTCTGCGGCTGTGGCTCTTCTT	240
QY	351 TTGTGGGGACGCTTTTGGCAGACCCAGATGTTTCTACTGATGAGCCCGCTGGC	410
Db	241 TTGTGGGGACGCTTTTGGCAGACCCAGATGTTTCTACTGATGAGCCCGCTGGC	300
QY	411 ACGTGTGATGACCTTTAAGCAGACGCCCTGGATGCTGCAATGAGCTGTGGGATC	470
Db	301 ACGTGTGATGACCTTTAAGCAGACGCCCTGGATGCTGCAATGAGCTGTGGGATC	360
QY	471 TGATACGGGCGTCTTCTTGTGGCAGTGAAGCTTTGATGCTTACATGGAAGCTGTG	530
Db	361 TGATACGGGCGTCTTCTTGTGGCAGTGAAGCTTTGATGCTTACATGGAAGCTGTG	420
QY	531 CCGGGAACAGTCCAGCCTTTTCAGTGGGAGAAACGCCGGCCCTGTGTTCTGCACTG	590
Db	421 CCGGGAACAGTCCAGCCTTTTCAGTGGGAGAAACGCCGGCCCTGTGTTCTGCACTG	480
QY	591 CCGTGTGATCATCCCAAGATGAATCATCCCGCGGCTCACTGAGGCTCCTGTGCA	650
Db	481 CCGTGTGATCATCCCAAGATGAATCATCCCGCGGCTCACTGAGGCTCCTGTGCA	540
QY	651 GTCAACAGCCCTTTGAGTGTGGAGAGGCTGCGCTCTCAAGCCACGTTGTTGCTCA	710
Db	541 GTCAACAGCCCTTTGAGTGTGGAGAGGCTGCGCTCTCAAGCCACGTTGTTGCTCA	600
QY	711 AAGAGGTGGCTTCTTCAACCTGCAAGTCCCTCTACCCGCTGTGAAGACCCCTCTCA	770
Db	601 AAGAGGTGGCTTCTTCAACCTGCAAGTCCCTCTACCCGCTGTGAAGACCCCTCTCA	660
QY	771 ACCTGATATCTGTGACACCTGTGTCCGGACCCCGGCGCTGTTCCTGTCCGAGACGA	830
Db	661 ACCTGATATCTGTGACACCTGTGTCCGGACCCCGGCGCTGTTCCTGTCCGAGACGA	720
QY	831 CAAAGGAAATCATGATGATGACAGTCCGATTTGTATGAGGGCAGATGAGCAAACTCA	890
Db	721 CAAAGGAAATCATGATGATGACAGTCCGATTTGTATGAGGGCAGATGAGCAAACTCA	780
QY	891 AGAAGGAGACCAACCTTACTATGTGATGCAAGTCACTGCCCCAAGCAGCTGAGATCT	950
Db	781 AGAAGGAGACCAACCTTACTATGTGATGCAAGTCACTGCCCCAAGCAGCTGAGATCT	840
QY	951 ACAAGACATCCAGTCTTGGCCCAAGGCCCTGCAGAAAGCTACCTGTGTGGCTATG	1010
Db	841 ACAAGACATCCAGTCTTGGCCCAAGGCCCTGCAGAAAGCTACCTGTGTGGCTATG	900
QY	1011 AGGACCTGGCTGAGGCCCTGTGGCCACAGCTTCCGGAATGTATGATTCGTGGATTGG	1070
Db	901 AGGACCTGGCTGAGGCCCTGTGGCCACAGCTTCCGGAATGTATGATTCGTGGATTGG	960
QY	1071 AATTTTGGCCCATCTTACAGCTGGGTGCATATCATCCCGAGGACGAGGATGGTG	1130
Db	961 AATTTTGGCCCATCTTACAGCTGGGTGCATATCATCCCGAGGACGAGGATGGTG	1020
QY	1131 ACCAGCCTTCCACACAATAATGACAGGATGCCCTTAATGTCTCCAGGCTTGGCCTGAT	1190
Db	1021 ACCAGCCTTCCACACAATAATGACAGGATGCCCTTAATGTCTCCAGGCTTGGCCTGAT	1080
QY	1191 CTTTGGCCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCGATGCCATGATTTGC	1250
Db	1081 CTTTGGCCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCGATGCCATGATTTGC	1140
QY	1251 TGGGCTACCGCCACGTCATCATCTGAACAGAAAGAAACCTGTGGGAAATCTTCGT	1310
Db	1141 TGGGCTACCGCCACGTCATCATCTGAACAGAAAGAAACCTGTGGGAAATCTTCGT	1200
QY	1311 CTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTCTGCCACT	1370
Db	1201 CTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTCTGCCACT	1260
QY	1371 GGTGTGAGGCTGAGTCACTTCTCTGAATGCTTGTGAGGCTTGGCTTACATCTCTAGGCT	1430
Db	1261 GGTGTGAGGCTGAGTCACTTCTCTGAATGCTTGTGAGGCTTGGCTTACATCTCTAGGCT	1320
QY	1431 TAACATCATGTCTGTGGATACACACTGAGTGTGATGTGTCCACACAGTGTCTCAAGAG	1490
Db	1321 TAACATCATGTCTGTGGATACACACTGAGTGTGATGTGTCCACACAGTGTCTCAAGAG	1380
QY	1491 AAGGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGGAACTTATGTAGCAG	1550
Db	1381 AAGGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGGAACTTATGTAGCAG	1440
QY	1551 CACATCCCAACAGTGAAGAGGATTTGCTCTTCTCTTGTGATCTCTCTGCTGGG	1610
Db	1441 CACATCCCAACAGTGAAGAGGATTTGCTCTTCTCTTGTGATCTCTCTGCTGGG	1500
QY	1611 CAGACTTCAAGAGACTTTGTGGCCTGGAGGCTATTTAAGCAGACACAGTATCATGTGAAT	1670
Db	1501 CAGACTTCAAGAGACTTTGTGGCCTGGAGGCTATTTAAGCAGACACAGTATCATGTGAAT	1560
QY	1671 TGATCCATTAACCTCCCTCTCCACATCTTGCCCAATGGGGAATGATCTTTACCAAGA	1730
Db	1561 TGATCCATTAACCTCCCTCTCCACATCTTGCCCAATGGGGAATGATCTTTACCAAGA	1620
QY	1731 GCTCACAGCATTTTCCACAGAGATGCAAAATTCAGAGCCCTTGAGTTCACAGTGAATTC	1790
Db	1621 GCTCACAGCATTTTCCACAGAGATGCAAAATTCAGAGCCCTTGAGTTCACAGTGAATTC	1680



```

|||||
Db 143604 TGGCGCTGCTTTTGCCCTATGAAAAGTTTCTCGACTTCAGAAAGCCTGTGGCATGCC 143663
OY 1241 ATGATTTGGCTGGGCTACCGGCACGCTGATCTGAAACAAGAACAGAACTGTGTG 1300
Db 143664 ATGAATTTGCTGGGCTACCGGCACGCTGATCTGAAACAAGAACAGAACTGTGTG 143723
OY 1301 GATCTTTCTGTACCTGGAGTGTCCCTAGCAAAATCCACTAAGAGGTTGAGAGGCTTT 1360
Db 143724 GATCTTTCTGTACCTGGAGTGTCCCTAGCAAAATCCACTAAGAGGTTGAGAGGCTTT 143783
OY 1361 GCTGCCACCTGTGTGACCTTCAGCTTTTCTGGAATGCTTCTGAGCCTTGCCATCAT 1420
Db 143784 GCTGCCACCTGTGTGACCTTCAGCTTTTCTGGAATGCTTCTGAGCCTTGCCATCAT 143843
OY 1421 CTCTAGGCTTTACTACATGCTGTGTGGGTATCCACACTAGTGTGATGTGTCCACAGCT 1480
Db 143844 CTCTAGGCTTTACTACATGCTGTGTGGGTATCCACACTAGTGTGATGTGTCCACAGCT 143903
OY 1481 GCTCAGAGAGAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGAGACCTT 1540
Db 143904 GCTCAGAGAGAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGAGACCTT 143963
OY 1541 ATGTAGAGAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTTCTTGATCTT 1600
Db 143964 ATGTAGAGAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTTCTTGATCTT 144023
OY 1601 CCTGTCTGGGAGAGACTTCAGAGACTTTGTGGCTGTGAGGCCCTATTAAACAGACAGAGTA 1660
Db 144024 CCTGTCTGGGAGAGACTTCAGAGACTTTGTGGCTGTGAGGCCCTATTAAACAGACAGAGTA 144083
OY 1661 TCAGTGGATTTGATCCATAAAGCTCCCTGTCCACTCTTGGCCAAATGGGGAATGATCTT 1720
Db 144084 TCAGTGGATTTGATCCATAAAGCTCCCTGTCCACTCTTGGCCAAATGGGGAATGATCTT 144143
OY 1721 TCACCAAGAGAGCTCACAGCACTTTTCCACAGAGATGCAAAATTCAGAGCCTTGGAGTTCC 1780
Db 144144 TCACCAAGAGAGCTCACAGCACTTTTCCACAGAGATGCAAAATTCAGAGCCTTGGAGTTCC 144203
OY 1781 CAGTGGATTCAGAGAGAGAGTGGAGACAGGTTGATGCTTATGAGCTTGACCAT 1840
Db 144204 CAGTGGATTCAGAGAGAGAGTGGAGACAGGTTGATGCTTATGAGCTTGACCAT 144263
OY 1841 CACAGCTATCGGTAATCAAGAAATATGAACAAAATCTGTGACAAAAGAGAACTCTTA 1900
Db 144264 CACAGCTATCGGTAATCAAGAAATATGAACAAAATCTGTGACAAAAGAGAACTCTTA 144323
OY 1901 AGTTCACAGAGGTGCTGGGCTGCACTTTGAATATCACTTCCCTGTGCAATTTTCCATCAC 1960
Db 144324 AGTTCACAGAGGTGCTGGGCTGCACTTTGAATATCACTTCCCTGTGCAATTTTCCATCAC 144383
OY 1961 ATAGAGAGCTTTGAGCCTGTGAAGCTGCATCTGTAATATCACTTCCCAAAATTAAG 2017
Db 144384 ATAGAGAGCTTTGAGCCTGTGAAGCTGCATCTGTAATATCACTTCCCAAAATTAAG 144440

```

```

RESULT 5
AF280088 1992 bp mRNA linear PRI 20-FEB-2001
LOCUS Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA,
DEFINITION complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1992)
AUTHORS Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
Chromosomal localization and genomic organization for the
galactose/ N-acetylgalactosamine/N-acetylgalactosamine

```

```

6-O-sulfotransferase gene family
Glycobiology 11 (1), 75-87 (2001)
MEDLINE
21096027
PUBMED
1181564
REFERENCE
2 (bases 1 to 1992)
AUTHORS Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES
location/Qualifiers
1..1992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1-q23.2"
/cell_type="high endothelial"
/tissue_type="tonsil"
1..1992
/gene="GST3"
122..1282
/gene="GST3"
/note="similar to Homo sapiens sequence encoded by GenBank
Accession Number AF131235"
/codon_start=1
/product="l-selectin ligand sulfotransferase GST-3"
/db_xref="GI:12060808"
/translation="MLPRKMLLLFLVSOMALLALFFHMSHNTSISMKRPRMH
VILSSNRGSSPVGQLEQHPDVEYILMPRAHWYTFKOSTAMLHNAVRLIAYF
LCMSVFDVIMYEPKRSLSLFWENSRALCSAPCDIIPQELIIPRAICRLCSQIP
FEVEKACRSYSHVLEVPFNLQSLVLPDPSLNILVHLDPRVPRVRSERIK
GDIMDSRIVMGHEQKLKEDQPYVMVQICQSLKTIKISLPKALOERYLLVRY
EDLARPVQVQTSMEYFVLEFLPHLOTWVNTTRGKGDHAPFTNADALNVQAW
RMSLPYKYSRLQKAGDAMNLGQVHVSDEQRLLDLISTVTPQHLH"
BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN
Query Match 86.4%; Score 1756; DB 9; Length 1992;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1856; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 160 CAAGGCTTCACCTTCACGCAATGCTACTGCTTAATAAATAGACTCTCTGCTTTCT 219
Db 100 CAAGGCTTCACCTTCACGCAATGCTACTGCTTAATAAATAGACTCTCTGCTTTCT 159
OY 220 GGTTCACAGATGGCCATCTTGGCTCTATCTTCCACATGTACAGCCACACATCAGCTC 279
Db 160 GGTTCACAGATGGCCATCTTGGCTCTATCTTCCACATGTACAGCCACACATCAGCTC 219
OY 280 CCGTCTATTAAGGACACAGCCCGGACATGACAGCTGTCTGTCTTCTCTGGGCGCTC 339
Db 220 CCGTCTATTAAGGACACAGCCCGGACATGACAGCTGTCTGTCTTCTCTGGGCGCTC 279
OY 340 TGGCTCTTCTTTGTGGGAGAGCTTTTGGGACAGCCAGATGTTTCTACTGTATGA 399
Db 280 TGGCTCTTCTTTGTGGGAGAGCTTTTGGGACAGCCAGATGTTTCTACTGTATGA 339
OY 400 GCCCGCTGGCAGTGTGATGACCTTCAAGCAGAGACCCGCTGATGTGACATGGC 459
Db 340 GCCCGCTGGCAGTGTGATGACCTTCAAGCAGAGACCCGCTGATGTGACATGGC 399
OY 460 TGTGGGAGATCTATATAGGGGCGGCTCTGTGTCGACATAGAGCTGTGGATGCTACAT 519
Db 400 TGTGGGAGATCTATATAGGGGCGGCTCTGTGTCGACATAGAGCTGTGGATGCTACAT 459
OY 520 GGAACCTGTGTCGGGAGACATGACCTCTTCAAGTGGGAGACAGCCGGGCGCTGTG 579
Db 460 GGAACCTGTGTCGGGAGACATGACCTCTTCAAGTGGGAGACAGCCGGGCGCTGTG 519
OY 580 TTCTGACCTGCTGTGACATCATCCACAGATGAATATGATCCCGGGGCTACTGACG 639
Db 520 TTCTGACCTGCTGTGACATCATCCACAGATGAATATGATCCCGGGGCTACTGACG 579

```

QY	640	GCTCTGTGACAGTAAACAGCCCTTTGAGGTGATGAGAGAGAGGCTGCCGCTCTACAGCCA	699
Db	580	GCTCTGTGACAGTAAACAGCCCTTTGAGGTGATGAGAGAGAGGCTGCCGCTCTACAGCCA	639
QY	700	CGTGTGTCTAAGAGAGGTGGGCTCTTCACACCTGCAGATCCCTCTACCCGCTGTGAAGA	759
Db	640	CGTGTGTCTAAGAGAGGTGGGCTCTCTTCACACCTGCAGATCCCTCTACCCGCTGTGAAGA	699
QY	760	CCCCCTCCCTACCTGCATATCTGTGCACCTGGTCCGGGACCCCCGGCGGTGTCCGTTTC	819
Db	700	CCCCCTCCCTACCTGCATATCTGTGCACCTGGTCCGGGACCCCCGGCGGTGTCCGTTTC	759
QY	820	CCGAGACGCGCAAGAGGAGATCTCATGATTGACATCGCATGTGTATGGGGAGCATGA	879
Db	760	CCGAGACGCGCAAGAGGAGATCTCATGATTGACATCGCATGTGTATGGGGAGCATGA	819
QY	880	GCAAAACCTCAAGAGAGGAGACCAACCCCTACTATGTATGATCAGGTGATCTGCCAAGCA	939
Db	820	GCAAAACCTCAAGAGAGGAGACCAACCCCTACTATGTATGATCAGGTGATCTGCCAAGCA	879
QY	940	GCTTGAGATCTACAGAGACATCCAGTCTCTGCCCAAGCCCTCAGAGACGCTACTGCT	999
Db	880	GCTTGAGATCTACAGAGACATCCAGTCTCTGCCCAAGCCCTCAGAGACGCTACTGCT	939
QY	1000	TGTGCGCATATAGACACCTGTGCTGAGCCCTGTGGCCACAGACTTCCCGAATGATCAAT	1059
Db	940	TGTGCGCATATAGACACCTGTGCTGAGCCCTGTGGCCACAGACTTCCCGAATGATCAAT	999
QY	1060	CGTGGGATTGGAATTTCTTGCCCATCTTCAGACCTGGGTGCATTAATCATCCCGAGCAA	1119
Db	1000	CGTGGGATTGGAATTTCTTGCCCATCTTCAGACCTGGGTGCATTAATCATCCCGAGCAA	1059
QY	1120	GGGCGATGGGTACACACCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC	1179
Db	1060	GGGCGATGGGTACACACCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC	1119
QY	1180	TTGGCGCTGTGTTTGCCCTATGAGAAAGTTTCTCGACTTCAGAAAGCTGTGGCATGC	1239
Db	1120	TTGGCGCTGTGTTTGCCCTATGAGAAAGTTTCTCGACTTCAGAAAGCTGTGGCATGC	1179
QY	1240	CATGATTTGCTGGGCTACCGCCACGTCAGATCTGAAACAGAAACAGAAACCTGTGCT	1299
Db	1180	CATGATTTGCTGGGCTACCGCCACGTCAGATCTGAAACAGAAACAGAAACCTGTGCT	1239
QY	1300	GGATCTTCTGTCTACTGAGCTGTCCCTGAGCAATCCACTAAGAGGGTTGAGAAGGCTT	1359
Db	1240	GGATCTTCTGTCTACTGAGCTGTCCCTGAGCAATCCACTAAGAGGGTTGAGAAGGCTT	1299
QY	1360	TGCTGCCACCTGGGTGACGCTCAGTCACTTTCTCTAATGCTTCGAGCCTTGCTTACA	1419
Db	1300	TGCTGCCACCTGGGTGACGCTCAGTCACTTTCTCTAATGCTTCGAGCCTTGCTTACA	1359
QY	1420	TCTCTGACCTTAATCATATGCTGTGGGTATCACACTGAGTGTGATTTGTCCACAG	1479
Db	1360	TCTCTGACCTTAATCATATGCTGTGGGTATCACACTGAGTGTGATTTGTCCACAG	1419
QY	1480	TGCTTCAGCAGAGAGACTTTTGTGTCCATGCTGTGTCTGAAAACAGACTGGGGAACT	1539
Db	1420	TGCTTCAGCAGAGAGACTTTTGTGTCCATGCTGTGTCTGAAAACAGACTGGGGAACT	1479
QY	1540	TATGTGAGCAGACATCCCAACAGTGAACAGGGTATGCTCTCTCTTTTGTATCT	1599
Db	1480	TATGTGAGCAGACATCCCAACAGTGAACAGGGTATGCTCTCTCTTTTGTATCT	1539
QY	1600	TCTGTCTGGGACAGCTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGACAGT	1659
Db	1540	TCTGTCTGGGACAGACTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGACAGT	1599
QY	1660	ATCAGTGAATTTGATTCATAAATCTCCCTGTCCACACTTTGGCCATGGGAAATGATCT	1719
Db	1600	ATCAGTGAATTTGATTCATAAATCTCCCTGTCCACACTTTGGCCATGGGAAATGATCT	1659

QY	1780	CCAGTGTATTCAGAGAGAGTGGGAAACAGGTTGGATGGCTACTTATCAGCTTGACCA	1839
Db	1720	CCAGTGTATTCAGAGAGAGTGGGAAACAGGTTGGATGCTACTTATCAGCTTGACCA	1779
QY	1840	TCACAGCTATGGGTATTCAGAAATATTAACCAAAATCTCTGCACAAAGCAGAGCTCTT	1899
Db	1780	TCACAGCTATGGGTATTCAGAAATATTAACCAAAATCTCTGCACAAAGCAGAGCTCTT	1839
QY	1900	AAGTTTCAGAGGCTCCTGGGCTGCATTTGATATCACTTCCCTCTGCATTTTCCATCA	1959
Db	1840	AAGTTTCAGAGGCTCCTGGGCTGCATTTGATATCACTTCCCTCTGCATTTTCCATCA	1899
QY	1960	CATGAGACTTTTACCTGTGACACTCCCATCTGTATATCTAAATTTCCCAATTAAG	2017
Db	1900	CATGAGACTTTTACCTGTGACACTCCCATCTGTATATCTAAATTTCCCAATTAAG	1957

RESULT 6			
LOCUS	AF149783	1333 bp	mRNA linear PRI 02-JUL-2001
DEFINITION	Homo sapiens L-selectin ligand sulfoltransferase mRNA, complete cds		
ACCESSION	AF149783		
VERSION	AF149783.1	GI:13897503	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Yeh,J.C., Hirakawa,N., Petrzyk,B., Nakayama,J., Ellis,L.G., Rabuka,D., Hindsaul,O., Matth,J.D., Lowe,J.B. and Fukuda,M.		
TITLE	Novel sulfated lymphocyte homing receptors and their control by a Core1 extension beta 1,3-N-acetylglucosaminyltransferase		
JOURNAL	Cell 105 (7), 957-969 (2001)		
MEDLINE	21332592		
PUBMED	11439191		
REFERENCE	2 (bases 1 to 1333)		
AUTHORS	Hirakawa,N. and Fukuda,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES			
source	1..1333		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
CDS	93..1253		
	/note="LST"		
	/codon_start=1		
	/product="L-selectin ligand sulfoltransferase"		
	/protein_id="AAK48417.1"		
	/db_xref="GI:13897504"		
	/translation="MLPKRMKLLFLVSOMAILAFPMYSHNTSLSKAOPRRMRHLCDSVFDAMERGPQRQSLSDWENSRLCSAPACDIIPOEITPRACRILCSQD FEVERAKRSYSHVLKEVFNFLQSLPLDPSLNHLVADPRAVRSRRT GMDLDSRIYMGHDEKLRKEDQPYVMVQISOLEIKTQSLPKLQERYLLVR EDLARAIVQATSRMYEFGVLEFLPHLQTVNHIITKGGGDHAFNTNADMLNVSQA RWSLPEKYSRLQACGDMLNGLGTHYVSEDEQRNLIDLILSTWVEQIH"		
BASE COUNT	286 a 393 c 336 g 318 t		
ORIGIN			
Query Match	59.6%;	Score 1211;	DB 9; Length 1333;
Best Local Similarity	99.9	Pred. No. 0;	
Matches 1261;	Conservative	0; Mismatches	1; Indels 0; Gaps 0;

QY	161	AAGGCTCCCACTTCACAGACATGCTACTGCTAAAAAATGAAGCTCCGCTGTTCTG	220
Db	72	AAGGCTTCACCTTCACAGACATGCTACTGCTAAAAAATGAAGCTCCGCTGTTCTG	131

[illegible]

QY	221	GTTCCTCCAGATGGCAATCTGGCTCTTAATCTTCACATCTACAGCACAACATCAAGCTCC	280
Db	132	GTTCCTCCAGATGGCAATCTGGCTCTTAATCTTCACATCTACAGCACAACATCAAGCTCC	191
QY	281	CTGTCTATGAAGGCACAGCCCGAGGCGCATGCAGCTGCTGTTCTCTTCCTGGCGCTCT	340
Db	192	CTGTCTATGAAGGCACAGCCCGAGGCGCATGCAGCTGCTGTTCTCTTCCTGGCGCTCT	251
QY	341	GGCTCTCTTTTGGGGGCGAGCTTTTGGGCGAGACCCAGATGTTTTTCTACTGATGAG	400
Db	252	GGCTCTCTTTTGGGGGCGAGCTTTTGGGCGAGACCCAGATGTTTTTCTACTGATGAG	311
QY	401	CCCGCCCTGGACAGCTGTGAGATGACCTTCAAGCAGAGACCGCCGCTGATGCTGCACATGAGT	460
Db	312	CCCGCCCTGGACAGCTGTGAGATGACCTTCAAGCAGAGACCGCCGCTGATGCTGCACATGAGT	371
QY	461	GTGGGGGATCTGATACGGGCGCTCTCTTGTGCGACATGAGCGCTTTGATGCTCTACTG	520
Db	372	GTGGGGGATCTGATACGGGCGCTCTCTTGTGCGACATGAGCGCTTTGATGCTCTACTG	431
QY	521	GAACCTGTCCTCCCGAGAGACAGTCCAGCCCTTCTTCAAGTGGGAAACAGCGCGGCGCTGT	580
Db	432	GAACCTGTCCTCCCGAGAGACAGTCCAGCCCTTCTTCAAGTGGGAAACAGCGCGGCGCTGT	491
QY	581	TCGTGACCTGCTGTGACATCATCTCCACAAGATGAATATATCCCCGGGCTCATCTGACGG	640
Db	492	TCGTGACCTGCTGTGACATCATCTCCACAAGATGAATATATCCCCGGGCTCATCTGACGG	551
QY	641	CTCCCTGTCAGTAAACAGCCCTTTGAGTGGTGGAGAAAGGCTGGCGGCTCAGCAGCCAC	700
Db	552	CTCCCTGTCAGTAAACAGCCCTTTGAGTGGTGGAGAAAGGCTGGCGGCTCAGCAGCCAC	611
QY	701	GTGTGCTCTAAGSAGGTGGCGCTTCTTCAACTGACCTGACCTCCCTTACCCGCTGCTGAAGAC	760
Db	612	GTGTGCTCTAAGSAGGTGGCGCTTCTTCAACTGACCTGACCTCCCTTACCCGCTGCTGAAGAC	671
QY	761	CCCTCCCTCAACCTGCAATATGATGAGACGATGTCGGGAGACCCCGGCGCTGTCGCTGCC	820
Db	672	CCCTCCCTCAACCTGCAATATGATGAGACGATGTCGGGAGACCCCGGCGCTGTCGCTGCC	731
QY	821	CGAGAAACGACAAAGAGGAGATCTCATGATTGAACATGCAATGATGATGAGGAGCAGCATGAG	880
Db	732	CGAGAAACGACAAAGAGGAGATCTCATGATTGAACATGCAATGATGATGAGGAGCAGCATGAG	791
QY	881	CAGAAATCTAAGAAGAGAGGACCAACCTCTATGTGATGACAGTATCTGCCAAAGCCAG	940
Db	792	CAGAAATCTAAGAAGAGAGGACCAACCTCTATGTGATGACAGTATCTGCCAAAGCCAG	851
QY	941	CTGAGATCTACAGACACATCCAGTCTCTTGSCCAAGAGCCCTCAGAGACGTAACCTGCTT	1000
Db	852	CTGAGATCTACAGACACATCCAGTCTCTTGSCCAAGAGCCCTCAGAGACGTAACCTGCTT	911
QY	1001	GTGGGCTATAGACACTGGCTGAGAGCCCTGTGGCCGACATTCGCCAATGTATGATATTC	1060
Db	912	GTGGGCTATAGACACTGGCTGAGAGCCCTGTGGCCGACATTCGCCAATGTATGATATTC	971
QY	1061	GTGGGATTTGAATTTCTGGCCCATTTCTAGACTGGGTGCATATACATCAACCGAGCAGAG	1120
Db	972	GTGGGATTTGAATTTCTGGCCCATTTCTAGACTGGGTGCATATACATCAACCGAGCAGAG	1031
QY	1121	GGCATGGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGCT	1180
Db	1032	GGCATGGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGCT	1091
QY	1181	TGGGCGCTGGCTTTGGCCCATGAAAGGTTTCTCGAGCTCACAAGAGCCTGTGGCGATGCC	1240
Db	1092	TGGGCGCTGGCTTTGGCCCATGAAAGGTTTCTCGAGCTCACAAGAGCCTGTGGCGATGCC	1151
QY	1241	ATGAAATTTGCTGGGCTACCGCCACAGTCAATCTGAAACAAGACAGAGAAACCTGTGTGTG	1300
Db	1152	ATGAAATTTGCTGGGCTACCGCCACAGTCAATCTGAAACAAGACAGAGAAACCTGTGTGTG	1211
QY	1301	GATCTTCTGTCTACCTGGAGCTGCTCCTTGAGCAATTCACATTAAGAGGCTTGAAGAGGCTTT	1360

	Db	1212	GATCTTCGTCTACTACGAGCAGTCCCTTGAGCAAAATCCACTAAGAGGGTTGGAGAAGCGTTTT	1271
QY	1361	GCTGCACCTGCTGGTGCACCCGACGACATTCTCTGATATGCTTCGAGCCCTTGCCATACAT	1420	
Db	1272	GCTGCACCTGCTGGTGCACCCGACGACATTCTCTGATATGCTTCGAGCCCTTGCCATACAT	1331	
QY	1421	CT 1422		
Db	1332	CT 1333		
RESULT 7				
LOCUS	AX381256	517 bp	DNA	Linear PAT 18-MAR-2002
DEFINITION	Sequence 194 from Patent WO0212280.			
ACCESSION	AX381256			
VERSION	AX381256.1	GI:19576075		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1	Pyle,R.A., Xu,J. and Secrist,H. Compositions and methods for the therapy and diagnosis of colon cancer Patent: WO 0212280-A 194 14-FEB-2002; CORIXA CORPORATION (US)		
AUTHORS	TITLE	Location/Qualifiers		
FEATURES	source	1..517		
	/organism="Homo sapiens"			
	/db_xref="taxon:9606			
BASE COUNT	121 a 130 c 124 g 142 t			
ORIGIN				
Query Match	24.9%; Score 505; DB 5; Length 517;			
Best Local Similarity	100.0%; Pred. NO. 5.6e-290; Indels 0; Gaps 0;			
Matches	505; Conservative 0; Mismatches 0;			
QY	1355	GGCTTTGCTGCCACCTGGCTGTCAAGCCTCAGNCACATTTCTCTGAATGCTTGTGAGCCTTGC	1414	
Db	13	GGCTTTGCTGCCACCTGGCTGTCAAGCCTCAGNCACATTTCTCTGAATGCTTGTGAGCCTTGC	72	
QY	1415	CTACATCTCTAGAGCCTTAACAATCTCTGTGGGTATACACTGAGTGTGATGTCTTC	1474	
Db	73	CTACATCTCTAGAGCCTTAACAATCTCTGTGGGTATACACTGAGTGTGATGTCTTC	132	
QY	1475	ACAGCTGCTCAAGCAGAGAGACTTTTGTGTCATGCTGTGTCTAGAAAACAGACTGGGG	1534	
Db	133	ACAGCTGCTCAAGCAGAGAGACTTTTGTGTCATGCTGTGTCTAGAAAACAGACTGGGG	192	
QY	1535	AACCTTATGTAGACACACATCCACCAGTAGAAAAAGSGATATGCTCTCTTCTTTCTT	1594	
Db	193	AACCTTATGTAGACACACATCCACCAGTAGAAAAAGSGATATGCTCTCTTCTTTCTT	252	
QY	1595	GATCTTCTGTCTGGGCGAGACTTCAGAAACATTTGGGCTTGAGGCGCTTAAGACAGAC	1654	
Db	253	GATCTTCTGTCTGGGCGAGACTTCAGAAACATTTGGGCTTGAGGCGCTTAATTAAGCACAC	312	
QY	1655	ACAGTATCAGTGAATGATCATTAACCTCCGTCGCACATCTTGGCCAAATGGGGAAATG	1714	
Db	313	ACAGTATCAGTGAATGATCATTAACCTCCGTCGCACATCTTGGCCAAATGGGGAAATG	372	
QY	1715	GATCTTACCAAAGAGCTCACCCAGCATTTTCCACAGAGATGCAAAATCTGAGCCCTTGG	1774	
Db	373	GATCTTACCAAAGAGCTCACCCAGCATTTTCCACAGAGATGCAAAATCTGAGCCCTTGG	432	
QY	1775	AGTTCCACAGTGTGANTCAAGAGAAAGTGGGAGAACAGTTTGATAGCCCTACTTATAGCCTT	1834	
Db	433	AGTTCCACAGTGTGANTCAAGAGAAAGTGGGAGAACAGTTTGATAGCCCTACTTATAGCCTT	492	
QY	1835	GACCATCAGAGCTATCGGTAATCAG 1859		



Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 678 AGGCTGCGGCTCTACAGCAGCAGTGCTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736  
|||||  
Db 488 AGGCTGCGGCTCTACAGCAGCAGTGCTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 546  
|||||

RESULT 11  
AF176838 2170 bp mRNA linear PRI 22-SEP-1999  
LOCUS Homo sapiens intestinal N-acetylglucosamine 6-O-sulfotransferase  
DEFINITION (1-1-GlcNAc-6-ST) mRNA, complete cds.  
ACCESSION AF176838  
VERSION AF176838.1 GI:5917705  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2170)  
AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.  
TITLE Cloning and characterization of a mammalian  
N-acetylglucosamine-6-sulfotransferase that is highly restricted to  
intestinal tissue  
JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)  
MEDLINE 99423499  
PUBMED 10491328  
REFERENCE 2 (bases 1 to 2170)  
AUTHORS Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.  
TITLE Direct Submission  
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,  
3401 Hillview Avenue, Palo Alto, CA 94304, USA

FEATURES  
source  
1..2170  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q23.1"  
/tissue\_type="Intestine"  
1..2170  
/gene="1-1-GlcNAc-6-ST"  
344..1516  
/gene="1-1-GlcNAc-6-ST"  
/note="possibly involved in biosynthesis of mucosal  
ligands for L-selection"  
/codon\_start=1  
/product="N-acetylglucosamine 6-O-sulfotransferase"  
/protein\_id="AAB5600.1"  
/db\_xref="GI:5917706"  
/translation="MMLPRESSKVTVLLLAQTCLLFLIISPGSSPAGGEDRVHV  
LVASWSRSGSFLGOLFQSDVFLMEPMHWITLSQSSATLHMAVDLMKSTFL  
CDMDVDPAYMPQRNLSAFPNNMATSRLCSPACSAFPRTISKQDVCKICTROFES  
LARACRSYSHVVLKEVRFNLQVLYPLSDPLNRIYHLVDPRAVLRSGRAAGDI  
LARDNGIVLTNGTWADPGRLVREVCVSHRTAEMATIKRPPLRGRLVREED  
LAREPLAEIRALYAFITGILTPLEENINHTTSGSGIETEFHTSSRRARVSGAM  
RHAPLPRKIRRVOLCAGALQDLGIRPVVISEDQRNLADLDVLPRLNGFTWASSTASH  
PRN"

BASE COUNT 449 a 682 c 612 g 427 t

Query Match 2.9%; Score 59; DB 9; Length 2170;  
Best Local Similarity 100.0%; Pred. No. 1e-22;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 678 AGGCTGCGGCTCTACAGCAGCAGTGCTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736  
|||||  
Db 834 AGGCTGCGGCTCTACAGCAGCAGTGCTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 892  
|||||

RESULT 12  
AF219990 2544 bp mRNA linear PRI 26-OCT-2000  
LOCUS Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase  
DEFINITION (CHST6) mRNA, complete cds.

ACCESSION AF219990  
VERSION AF219990.1 GI:11023145  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2544)  
AUTHORS Akama,T.O., Nishida,K., Nakayama,A., Watanabe,H., Fujiwara,T.,  
Nakanura,T., Dots,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,  
Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.  
TITLE Macular corneal dystrophy type I and type II are caused by distinct  
mutations in a new sulfotransferase gene  
JOURNAL Nat. Genet. 26 (2), 237-241 (2000)  
MEDLINE 20472330  
PUBMED 11017086  
REFERENCE 2 (bases 1 to 2544)  
AUTHORS Akama,T.O. and Fukuda,M.N.  
TITLE Direct Submission  
JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham  
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA  
location/Qualifiers  
1..2544  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q22"  
1..2544  
/gene="CHST6"  
/note="responsible for macular corneal dystrophy"  
693..1880  
/gene="CHST6"  
/function="transfers sulfate to 6-OH of GlcNAc in keratan  
sulfate"  
/note="C-GlcNAc6ST"  
/product="corneal  
N-acetylglucosamine-6-O-sulfotransferase"  
/protein\_id="AAG26325.1"  
/db\_xref="GI:11023146"  
/translation="MMLPVRSSTAVTALLLAQTFLLELVSRPGSSPAGGEARVHL  
VLWSRSGSFEVGLFNOHPDVFLMEPMHWITLSQSSATLHMAVDLVRVFLC  
DMDVDPAYMPQRNLSDFQWAVSRALCSPACSAFPRTISSEAVKPLARQSFIL  
AREACRSYSHVVLKEVRFNLQVLYPLSDPLNRIYHLVDPRAVLRSGRAAGDI  
ARDNGIVLTNGTWADPGRLVREVCVSHRTAEMATIKRPPLRGRLVREED  
LAREPLAEIRALYAFITGILTPLEENINHTTSGSGIETEFHTSSRRARVSGAM  
RHAPLPRKIRRVOLCAGALQDLGIRPVVISEDQRNLADLDVLPRLNGFTWASSTASH  
PRN"

BASE COUNT 460 a 799 c 733 g 552 t

Query Match 2.9%; Score 59; DB 9; Length 2544;  
Best Local Similarity 100.0%; Pred. No. 1e-22;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 678 AGGCTGCGGCTCTACAGCAGCAGTGCTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736  
|||||  
Db 1180 AGGCTGCGGCTCTACAGCAGCAGTGCTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 1238  
|||||

RESULT 13  
AF246718 3278 bp mRNA linear PRI 31-OCT-2000  
LOCUS Homo sapiens intestinal GlcNAc-6-sulfotransferase (CHST5) mRNA,  
DEFINITION complete cds, alternatively spliced.  
ACCESSION AF246718  
VERSION AF246718.1 GI:11055254  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3278)



REFERENCE	TITLE	FEATURES
JOURNAL MEBLINE PUBMED 1181864	Chromosomal localization and genomic organization for the galactose/ N-acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene family Glycobiology 11 (1), 75-87 (2001)	
AUTHORS	2 (bases 1 to 3786) Hemmerich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and Rosen, S.D.	
JOURNAL	Direct Submission Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA	
SOURCE	Location/Qualifiers 1..3786	
gene	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="16q23.1" 1..3786	
CDS	/gene="GST4beta" /note="clusters with the intestinal N-acetylglucosamine 6-O-sulfotransferase (GSTAlpha) gene in GenBank Accession Numbers AF176838 and AF176839" 189..1376 /gene="GST4beta" /note="similar to N-acetylglucosamine 6-O-sulfotransferase GST-4alpha" /codon_start=1 /product="N-acetylglucosamine 6-O-sulfotransferase GST-4beta" /protein_id="AA048244.1" /db_xref="GI:12060804" /translation="MMLPRVSTAVTALLAQTFLLFLVSRPPSPAGGEARVHVLVSRRSGSFVQGLERQHDVEFLMEPAHWTTISQSAATLHMARQIVASVFLDMQVDFALPMPRNLSDFQWASRALCSPACAPRGASISEAVCKPLCARQSFLLARECRSHSVLVKEVRFENLOVLIPLISPDALNIRVHVLRPRVAVRERQTKARLARDGIVLTGNGTINVEADPEGLRVYREYCRSHVTADEATIKPPLRGRIYLPFDLAREDELRALVATGSLTLPQLEAMVHNTTHSGGPEAREAKRTSRNLVNSQARHAPFAKIRRVQELCAQALQILGYRPSYSEDEQRNLALDLVLRGLNGFTWASSTASHPRN"	
polyA_signal	3106..3111 /gene="GST4beta" /note="alternative signal"	
polyA_signal	3734..3739 /gene="GST4beta"	
BASE COUNT	848 a 1030 c 1104 g 804 t	
ORIGIN		
Query Match	2.9% Score 59; DB 9; Length 3786;	
Best Local Similarity	100.0%; Pexd. No. 1.1e-22;	
Matches	59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	678 AGGCTGCGCGCTCTACAGCCAGCGTGTGCTCAAGAGAGTGCGCTCTCAACCTGACAG 736                               676 AGGCTGCGCGCTCTACAGCCAGCGTGTGCTCAAGAGAGTGCGCTCTCAACCTGACAG 734	
RESULT 15		
LOCUS	AF219991	
DEFINITION	Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase (CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) genes, complete cds.	
ACCESSION	AF219991	
VERSION	AF219991.1	
KEYWORDS	GI:11023147	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 71503) Akama, T.O., Nishida, K., Nakayama, J., Watanabe, H., Fujiwara, T., Nakamura, T., Doita, A., Kawasaki, S., Inoue, Y., Maeda, N., Yamamoto, S.,	





/rpt\_family="Alu"  
 /rpt\_type=dispersed  
 repeat\_region 21725..22207  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 repeat\_region 22008..22040  
 /rpt\_family="Simple\_repeat"  
 /rpt\_type=tandem  
 repeat\_region 22041..22313  
 /rpt\_family="MER4-group"  
 /rpt\_type=dispersed  
 repeat\_region 22325..22593  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 repeat\_region 22594..22949  
 /rpt\_family="MER4-group"  
 /rpt\_type=dispersed  
 repeat\_region 23116..23221  
 /rpt\_family="Simple\_repeat"  
 /rpt\_type=tandem  
 repeat\_region 23344..23400  
 /rpt\_family="MaLR"  
 /rpt\_type=dispersed  
 repeat\_region 23433..23458  
 /rpt\_family="Simple\_repeat"  
 /rpt\_type=tandem  
 repeat\_region complement(23457..23720)  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 repeat\_region complement(23789..24140)  
 /rpt\_family="MER21-group"  
 /rpt\_type=dispersed  
 repeat\_region 24141..24287  
 /rpt\_family="MaLR"  
 /rpt\_type=dispersed  
 repeat\_region complement(24308..24420)  
 /rpt\_family="MaLR"  
 /rpt\_type=dispersed  
 repeat\_region complement(24467..24528)  
 /rpt\_family="MER1\_type"  
 /rpt\_type=dispersed

Query Match 2.9%; Score 59; DB 9; Length 71503;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCCGCTCTACAGCAGCGTGCTGTCAAGAGGCGCTTCTTCAACCTGCAG 736  
 |||||  
 Db 19520 AGGCGTCCGCTCTACAGCAGCGTGCTGTCTCAAGAGGCGCTTCTTCAACCTGCAG 19578

RESULT 16  
 AC009163/ 157337 bp DNA linear PRI 02-NOV-2001  
 LOCUS Homo sapiens chromosome 16 clone RP11-77K12, complete sequence.  
 DEFINITION  
 AC009163  
 VERSION AC009163.5 GI:16596526  
 KEYWORDS  
 HTG.  
 SOURCE Homo sapiens.  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 157337)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 16  
 JOURNAL  
 2 (bases 1 to 157337)  
 REFERENCE  
 2 (bases 1 to 157337)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL  
 Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 REFERENCE  
 3 (bases 1 to 157337)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission

JOURNAL  
 Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 On Nov 2, 2001 this sequence version replaced gi:17689976.  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----

FEATURES  
 source  
 1..157337  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-77K12"

BASE COUNT 45295 a 36632 c 36308 g 39102 t  
 ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 157337;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-22;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCCGCTCTACAGCAGCGTGCTGTCAAGAGGCGCTTCTTCAACCTGCAG 736  
 |||||  
 Db 144095 AGGCGTCCGCTCTACAGCAGCGTGCTGTCAAGAGGCGCTTCTTCAACCTGCAG 144037

RESULT 17  
 AC025287 194832 bp DNA linear PRI 26-JAN-2002  
 LOCUS Homo sapiens chromosome 16 clone RP11-490818, complete sequence.  
 DEFINITION  
 AC025287  
 VERSION AC025287.8 GI:18376863  
 KEYWORDS  
 HTG.  
 SOURCE Homo sapiens.  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 194832)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL  
 2 (bases 1 to 194832)  
 REFERENCE  
 2 (bases 1 to 194832)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL  
 Submitted (21-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 REFERENCE  
 4 (bases 1 to 194832)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL  
 Submitted (26-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 On Jan 26, 2002 this sequence version replaced gi:17976465.  
 Draft Sequence Produced by DOE Joint Genome Institute

```

www.jgi.doe.gov/Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.6.
Location/Qualifiers
1. 194832
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-490B18"
BASE COUNT 53784 a 43240 c 43815 g 53993 t
ORIGIN
Query Match 2.9% Score 59; DB 9; Length 194832;
Best Local Similarity 100.0%; Pred. No. 1.3e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 678 AGGCGTCGCCCTCTACAGCCAGCGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736
|||||
Db 188038 AGGCGTCGCCCTCTACAGCCAGCGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGCAG 188096
|||||
RESULT 18 208185 bp DNA linear HTG 25-APR-2001
AC009105
LOCUS AC009105
DEFINITION Homo sapiens chromosome 16 clone RP11-455E15, WORKING DRAFT
SEQUENCE 17 unordered pieces.
AC009105.7 GI:13786304
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208185)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 208185)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:1689928.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RP11-455E15
-----
Summary Statistics
Consensus quality: 180040 bases at least Q40
Consensus quality: 182095 bases at least Q30
Consensus quality: 195726 bases at least Q20
Estimated insert size: 195650; agarose-1p estimation
Estimated insert size: 206585; sum-of-coverage estimation
Quality coverage: 6.57 in Q20 bases; agarose-1p estimation
Quality coverage: 6.23 in Q20 bases; sum-of-coverage estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1062: contig of 1062 bp in length
* 1063 1162: gap of unknown bp in length
* 1163 235: contig of 1169 bp in length

```

QY 884 AACTCAAGAGAGACCAACCTACTATGTGATGC 920  
 AF131236  
 LOCUS AF131236  
 DEFINITION Mus musculus N-acetylglucosamine 6-O-sulfotransferase gene, complete cds.  
 ACCESSION AF131236  
 VERSION AF131236.1 GI:4927115  
 KEYWORDS Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1926)  
 AUTHORS Bistup, A., Bhakta, S., Lee, J.-K., Belov, Y. Y., Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Rosen, S. D. and Hemmerlich, S.  
 TITLE Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin  
 JOURNAL J. Cell Biol. 145 (4), 899-910 (1999)  
 MEDLINE 99264336  
 PUBMED 10330415  
 REFERENCE 2 (bases 1 to 1926)  
 AUTHORS Bistup, A., Tangemann, K., Bhakta, S., Lee, J.-K., Belov, Y. Y., Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Rosen, S. D. and Hemmerlich, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA  
 FEATURES  
 Source  
 1. 1926  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /cell\_type="high endothelial"  
 /tissue\_type="tonsil"  
 <405..>1571  
 /product="N-acetylglucosamine 6-O-sulfotransferase"  
 405..1571  
 /function="Biosynthesis of endothelial ligands for L-selectin"  
 /note="HEC-GLCNAc-6-Sr"  
 /codon\_start=1  
 /product="N-acetylglucosamine 6-O-sulfotransferase"  
 /protein\_id="AAD33016.1"  
 /db\_xref="GI:4927116"  
 /translation="MMLKRGRLMFLGSOIVVALLFIMSVHRHLSQRESRPVAH  
 LVLSWRSGSFVGOLEFGHPDVEYLMPEAMHWMTFTSTAMKIHMAVRDLRSVFL  
 CMSSYFDAYMNPGRKOSLFWMOBSRALCSVPCDFPAHRTISSPKCKLIGQGF  
 DMVEKACRSHGFEVLEKVRFFSLQALPLPLDPSLNHVLVLRPRAVPSREHTTI  
 ELMBDSHVLVGHLETLKEEDQPYAMKTKSSQVDIVKAIQTLFEALQQRFLRYE  
 ELVRAFLAQTRLYKRVGLDELPLQATQVYNTVRKGMQGHAFHNANANAVSQMR  
 WSLPEYKVSQLDACGFAMDGLGLQVRSQCGNLSLILSSHILQVREG"  
 BASE COUNT 426 a 520 c 484 g 496 t  
 ORIGIN  
 Query Match 1.3%; Score 27; DB 10; Length 1926;  
 Best Local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 494 GACATGAGCGCTTTGATGCTACATG 520  
 AF109155  
 LOCUS AF109155  
 DEFINITION Mus musculus L-selectin ligand sulfotransferase mRNA, complete cds.

ACCESSION AF109155  
 VERSION AF109155.1 GI:5596405  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2201)  
 AUTHORS Hiraoka, N., Petryniak, B., Nakayama, J., Tsuboi, S., Suzuki, M., Yeh, J. C., Izawa, D., Tanaka, T., Miyasaka, M., Lowe, J. B. and Fukuda, M.  
 TITLE A novel, high endothelial venule-specific sulfotransferase expresses 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34  
 JOURNAL Immunology 11 (1), 79-89 (1999)  
 MEDLINE 99361934  
 PUBMED 10435581  
 REFERENCE 2 (bases 1 to 2201)  
 AUTHORS Hiraoka, N. and Fukuda, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-NOV-1998) Glycobiology, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA  
 FEATURES  
 Source  
 1. 2201  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /protein\_id="AAB45579.1"  
 /translation="directs expression of 6-sulfo sialyl Lewis X, an L-selectin ligand displayed by CD34"  
 /note="155r"  
 /codon\_start=1  
 /product="L-selectin ligand sulfotransferase"  
 /protein\_id="AAB45579.1"  
 /db\_xref="GI:5596406"  
 /translation="MMLKRGRLMFLGSOIVVALLFIMSVHRHLSQRESRPVAH  
 LVLSWRSGSFVGOLEFGHPDVEYLMPEAMHWMTFTSTAMKIHMAVRDLRSVFL  
 CMSSYFDAYMNPGRKOSLFWMOBSRALCSVPCDFPAHRTISSPKCKLIGQGF  
 DMVEKACRSHGFEVLEKVRFFSLQALPLPLDPSLNHVLVLRPRAVPSREHTTI  
 ELMBDSHVLVGHLETLKEEDQPYAMKTKSSQVDIVKAIQTLFEALQQRFLRYE  
 ELVRAFLAQTRLYKRVGLDELPLQATQVYNTVRKGMQGHAFHNANANAVSQMR  
 WSLPEYKVSQLDACGFAMDGLGLQVRSQCGNLSLILSSHILQVREG"  
 BASE COUNT 537 a 629 c 534 g 501 t  
 ORIGIN  
 Query Match 1.3%; Score 27; DB 10; Length 2201;  
 Best Local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 494 GACATGAGCGCTTTGATGCTACATG 520  
 AF302109  
 LOCUS AF302109  
 DEFINITION Homo sapiens ps20 WAF-type four-disulfide core domain protein mRNA, complete cds.  
 ACCESSION AF302109  
 VERSION AF302109.1 GI:10198226  
 KEYWORDS Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1398)  
 AUTHORS Ong, C. K., Ng, C. Y., Lim, K. B., Chan, T. W. M. G. and Huynh, H.  
 TITLE Molecular Cloning and Characterization of the Human PS20 protein in Human Uterus  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1398)  
 AUTHORS Ong, C. K., Ng, C. Y., Lim, K. B., Chan, T. W. M. G. and Huynh, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-SEP-2000) Cellular And Molecular Research, National

Cancer Centre of Singapore, No.11 Hospital Drive, Singapore 169610,  
Singapore

FEATURES  
Source Location/Qualifiers

CDS  
1..1398  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/tissue\_type="uterus"  
178..684  
/product="ps20 WAP-type four-disulfide core domain  
protein"  
/protein\_id="FAGI5263.1"  
/db\_xref="GI:10198227"  
/translation="MPLTGVPGSGCRQITRALCLLLILHAGSAKNIMKRALPARIA  
EKSRAEAGAGGRRQPRADRCPPRTLPFGACQARCGADSECPHRCYNGCAY  
ACLEAVPPVLDMLVQPKPRWLGNGMLDGPGEVLQAEACSTEDGAPLCPSCY  
ECHILBR"

BASE COUNT 335 a 407 c 400 g 256 t  
ORIGIN

Query Match 1.28; Score 24; DB 9; Length 1398;  
Best Local Similarity 100.0%; Pred. No. 0.094;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2009 CCAATAGAAAAA 2032  
Db 1371 CCAATAGAAAAA 1394

RESULT 23  
AC012281/c 79598 bp DNA linear HTG 22-OCT-1999  
LOCUS Homo sapiens chromosome 3 clone 40\_E\_23 map 3, LOW-PASS SEQUENCE  
DEFINITION SAMPLING.

AC012281  
AC012281.1 GI:6094608  
VERSION HTG; HTGS\_PHASED.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM

REFERENCE  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 79598)  
JOURNAL Bitten,B., Linton,L., Nusbaum,C. and Lander,E.  
REFERENCE Homo sapiens chromosome 3, clone 40\_E\_23  
AUTHORS Unpublished  
2 (bases 1 to 79598)  
Bitten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,D., Barna,N., Beckert,R., Boguslavsky,L., Bouknight,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hages,B., Hearford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,  
Lehocky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,  
McMan,P., McGuck,A., McKernan,K., McLaughlin,J., Meddum,J.,  
Morrow,J., Naylor,D., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html.

\* NOTE: This record contains 84 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 952: contig of 952 bp in length  
953 1890: contig of 938 bp in length  
1891 2864: gap of unknown length  
2865 3823: contig of 974 bp in length  
3824 4788: gap of unknown length  
4789 5761: contig of 965 bp in length  
5762 6656: gap of unknown length  
6657 7647: contig of 973 bp in length  
7648 8615: gap of unknown length  
8616 9594: contig of 991 bp in length  
9595 10541: gap of unknown length  
10542 11508: contig of 947 bp in length  
11509 12454: gap of unknown length  
12455 13400: contig of 967 bp in length  
13401 14343: gap of unknown length  
14344 15284: contig of 943 bp in length  
15285 16252: gap of unknown length  
16253 17230: contig of 941 bp in length  
17231 18195: gap of unknown length  
18196 19132: contig of 968 bp in length  
19133 20087: gap of unknown length  
20088 20897: contig of 937 bp in length  
20898 21802: gap of unknown length  
21803 22757: contig of 955 bp in length  
22758 23708: gap of unknown length  
23709 24654: contig of 951 bp in length  
24655 25603: gap of unknown length  
25604 26542: contig of 946 bp in length  
26543 27466: gap of unknown length  
27467 28441: contig of 924 bp in length  
28442 29397: gap of unknown length  
29398 30346: contig of 975 bp in length  
30347 31299: gap of unknown length  
31299 32254: contig of 956 bp in length  
32254 33300: gap of unknown length  
33300 33254: contig of 949 bp in length  
33254 33254: contig of 953 bp in length  
33254 33254: contig of 955 bp in length  
33254 33254: gap of unknown length

```

* 32255 33204: contig of 950 bp in length
* gap of unknown length
* 33205 34156: contig of 952 bp in length
* gap of unknown length
* 34157 35051: contig of 895 bp in length
* gap of unknown length
* 35052 36006: contig of 955 bp in length
* gap of unknown length
* 36007 36955: contig of 949 bp in length
* gap of unknown length
* 36956 37899: contig of 944 bp in length
* gap of unknown length
* 37900 38880: contig of 981 bp in length
* gap of unknown length
* 38881 39843: contig of 963 bp in length
* gap of unknown length
* 39844 40789: contig of 946 bp in length
* gap of unknown length
* 40790 41696: contig of 907 bp in length
* gap of unknown length
* 41697 42487: contig of 791 bp in length
* gap of unknown length
* 42488 43448: contig of 961 bp in length
* gap of unknown length
* 43449 44434: contig of 986 bp in length
* gap of unknown length
* 44435 45348: contig of 914 bp in length
* gap of unknown length
* 45349 46306: contig of 958 bp in length
* gap of unknown length
* 46307 47248: contig of 942 bp in length
* gap of unknown length
* 47249 48181: contig of 933 bp in length
* gap of unknown length
* 48182 49144: contig of 963 bp in length
* gap of unknown length
* 49145 50101: contig of 957 bp in length
* gap of unknown length
* 50102 51049: contig of 948 bp in length
* gap of unknown length
* 51050 51982: contig of 933 bp in length
* gap of unknown length
* 51983 52928: contig of 946 bp in length
* gap of unknown length
* 52929 53867: contig of 939 bp in length
* gap of unknown length
* 53868 54843: contig of 976 bp in length
* gap of unknown length
* 54844 55755: contig of 912 bp in length
* gap of unknown length
* 55756 56703: contig of 948 bp in length
* gap of unknown length
* 56704 57660: contig of 957 bp in length
* gap of unknown length
* 57661 58593: contig of 933 bp in length
* gap of unknown length
* 58594 59547: contig of 954 bp in length
* gap of unknown length
* 59548 60496: contig of 949 bp in length
* gap of unknown length
* 60497 61450: contig of 954 bp in length
* gap of unknown length
* 61451 62409: contig of 959 bp in length
* gap of unknown length
* 62410 63368: contig of 959 bp in length
* gap of unknown length
* 63369 64320: contig of 952 bp in length
* gap of unknown length
* 64321 65317: contig of 997 bp in length
* gap of unknown length
* 65318 66295: contig of 978 bp in length
* gap of unknown length
* 66296 67201: contig of 906 bp in length

```

```

Query Match      1.2%; Score 24; DB 2; Length 79598;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATTAGAAAAAAAAAAAAA 2031
Db 60724 CCCAATTAGAAAAAAAAAAAAA 60701

RESULT 24
AC095651/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC095651 99395 bp DNA linear HTG 11-JUL-2002
Rattus norvegicus clone CH230-8E2. *** SEQUENCING IN PROGRESS ***
44 unordered pieces.
AC095651.3 GI:21722647
HTG: HTGS PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 99395)
Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
Alstrooks S.L., Amaralunge H.C., Are J.R., Ayale M., Banks T.,
Barbataia V., Benton J., Bimaye K., Blanchard K., Bonnin D.,
Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burkh P., Burkett C., Burrell K.L., Byrd N.C.,
Carton T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhury I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Devila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
DeLaney K.R., Delgado O., Denn A.L., Ding T., Dinh H.H.,
Douthwaite K.J., Diaper H., Dugan-Rocha S., Durbin K.J.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escoto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Franz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Hawlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,
Hornli F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlsen E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kurishi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtenarge O., Lieu C., Liu J., Liu W., Louisge H.,
Lozada R.T., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapa P., Martin R., Martinale A., Martinez E.,
Massey E., Mawhinney E., McLeod M.P., Meador M., Mel G., Metzker M.,
Miner G., Miner Z., Mitchell T., Mohabati K., Morgan M., Morris S.,
Mosier M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokkenko S., Oguni M., Okunolu G.,
Orangnye N., Oyiedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rivers M., Rojas A., Rojupokan I., Rolfe M., Ruiz S., Savery G.,
Scherrer S., Scott G., Shen H., Shoshitari N., Sisson I.,
Sodergren E., Sotaike T., Sparks A., Stanley H., Stone H.,
Sutton A., Syatek A., Tabor P., Tameis A., Tameis K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalón D., Vinson R., Wang Q.,

```

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 99395)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-FEB-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 99395)  
 Worley, K.C.  
 Direct Submission  
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 10, 2002 this sequence version replaced g1:17942210.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc.help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GCXS  
 Center clone name: CH230-8E2  
 ----- Summary Statistics  
 Sequencing vector: plasmid  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap, version 0.990329  
 Consensus quality: 66825 bases at least Q40  
 Consensus quality: 71670 bases at least Q30  
 Consensus quality: 73911 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NAME: This is a 'working draft' sequence. It currently  
 \* consists of 44 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1425: contig of 1425 bp in length  
 \* 1426 1525: gap of unknown length  
 \* 1526 2919: contig of 1394 bp in length  
 \* 2920 3019: gap of unknown length  
 \* 3020 4525: contig of 1506 bp in length  
 \* 4526 4625: gap of unknown length  
 \* 4626 5831: contig of 1206 bp in length  
 \* 5832 5931: gap of unknown length  
 \* 5932 6935: contig of 1004 bp in length  
 \* 6936 7035: gap of unknown length  
 \* 7036 8283: contig of 1248 bp in length  
 \* 8284 8383: gap of unknown length  
 \* 8384 9805: contig of 1422 bp in length  
 \* 9806 9905: gap of unknown length  
 \* 9906 11125: contig of 1221 bp in length  
 \* 11127 11226: gap of unknown length  
 \* 1127 12713: contig of 1487 bp in length  
 \* 12714 12813: gap of unknown length  
 \* 12814 13930: contig of 1117 bp in length  
 \* 13931 14030: gap of unknown length  
 \* 14031 15456: contig of 1426 bp in length  
 \* 15457 15556: gap of unknown length  
 \* 1557 16890: contig of 1334 bp in length  
 \* 16891 16990: gap of unknown length  
 \* 16991 18008: contig of 1018 bp in length  
 \* 18009 18108: gap of unknown length  
 \* 18109 19518: contig of 1410 bp in length  
 \* 19519 19618: gap of unknown length

19619 20891: contig of 1273 bp in length  
 20892 20991: gap of unknown length  
 20992 23151: contig of 2160 bp in length  
 23152 23251: gap of unknown length  
 23252 24300: contig of 1049 bp in length  
 24301 24400: gap of unknown length  
 24401 25937: contig of 1537 bp in length  
 25938 26037: gap of unknown length  
 26038 27412: contig of 1375 bp in length  
 27413 27512: gap of unknown length  
 27513 28754: contig of 1222 bp in length  
 28735 28834: gap of unknown length  
 28835 29956: contig of 1122 bp in length  
 29957 30056: gap of unknown length  
 30057 31347: contig of 1291 bp in length  
 31348 31447: gap of unknown length  
 31448 32624: contig of 1177 bp in length  
 32625 32724: gap of unknown length  
 32725 34035: contig of 1311 bp in length  
 34036 34135: gap of unknown length  
 34136 35432: contig of 1297 bp in length  
 35433 35532: gap of unknown length  
 35533 37135: contig of 1603 bp in length  
 37136 37235: gap of unknown length  
 37236 39141: contig of 1906 bp in length  
 39142 39241: gap of unknown length  
 39242 40831: contig of 1590 bp in length  
 40832 40931: gap of unknown length  
 40932 43324: contig of 2393 bp in length  
 43325 43424: gap of unknown length  
 43425 45280: contig of 1866 bp in length  
 45281 45390: gap of unknown length  
 45391 47689: contig of 2279 bp in length  
 47690 47769: gap of unknown length  
 47770 49684: contig of 1895 bp in length  
 49685 49764: gap of unknown length  
 49765 52591: contig of 2827 bp in length  
 52592 52691: gap of unknown length  
 52692 55345: contig of 2654 bp in length  
 55346 55445: gap of unknown length  
 55446 58258: contig of 2813 bp in length  
 58259 58358: gap of unknown length  
 58359 60686: contig of 2338 bp in length  
 60687 60786: gap of unknown length  
 60787 63505: contig of 2719 bp in length  
 63506 63605: gap of unknown length  
 63606 66610: contig of 3005 bp in length  
 66611 72030: gap of unknown length  
 72031 72130: contig of 5320 bp in length  
 72131 75957: gap of unknown length  
 75958 76057: gap of unknown length  
 76059 79361: contig of 3304 bp in length  
 79362 79461: gap of unknown length  
 79462 84332: contig of 4861 bp in length  
 84333 84482: gap of unknown length  
 84483 91488: contig of 7066 bp in length  
 91489 91588: gap of unknown length  
 91589 93935: contig of 7807 bp in length.  
 Location/Qualifiers  
 1. 99395  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-8E2"  
 BASE COUNT 26409 a 18452 c 18712 g 27333 t 8489 others  
 ORIGIN  
 Query Match 1.2%: Score 24; DB 2; Length 99395;  
 Best Local Similarity 100.0%: Pred. No. 0.12;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 591 CCTGTGACATCATCCACCAAGAGC 614  
 {}

Db 45144 CCTGTGACATCATCCACAGATG 45121

RESULT 25  
AC116944/c  
LOCUS  
DEFINITION  
AC116944 131969 bp DNA linear HTG 14-JUL-2002  
Tetraodon nigroviridis clone GSTNB-15B16, WORKING DRAFT SEQUENCE, 3  
ordered pieces.

ACCESSION  
AC116944.2 GI:21389263  
VERSION  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
KEYWORDS  
Tetraodon nigroviridis.  
SOURCE  
Tetraodon nigroviridis  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

REFERENCE  
AUTHORS  
Ahter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,  
Brooks, S., Dietrich, N.L., Granter, S., Guan, X., Gupta, J.,  
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Latic, P.,  
Lee, H.S.-O., Legaspi, R., Maduro, O.L., Maduro, V.B.,  
Margulies, E.H., Mastello, C., Maskeri, B., Mastrian, S.D.,  
McCloskey, J.C., McDowell, J., Paquinigan, C., Pearson, R.,  
Portnoy, M.E., Prasad, A., Schneider, M.G., Stantrop, S., Thomas, J.W.,  
Thomas, P.J., Touchman, J.W., Tsurgou, C., Vogt, J.L., Walker, M.A.,  
Wetherby, K.D., Wiggin, L., Young, A., Zhang, L.H. and Green, E.D.  
NISC Comparative Sequencing Initiative

TITLE  
Unpublished

REFERENCE  
2 (bases 1 to 131969)

AUTHORS  
Green, E.D.

JOURNAL  
Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717  
Government Circle, Gaithersburg, MD 20877, USA

REFERENCE  
3 (bases 1 to 131969)

AUTHORS  
Green, E.D.

JOURNAL  
Submitted (14-JUL-2002) NIH Intramural Sequencing Center, 8717  
Government Circle, Gaithersburg, MD 20877, USA

COMMENT  
On Jun 11, 2002 this sequence version replaced gi:19909418.  
----- Genome Center

Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: nisc-zoo@nigrl.nih.gov  
----- Project Information  
Center project name: cvt  
Center clone name: 015B16

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 131566 bases at least Q40  
Consensus quality: 131709 bases at least Q30  
Consensus quality: 131757 bases at least Q20  
Insert size: 135000; agarose-ff  
Insert size: 131769; sum-of-contigs  
Quality coverage: 10.24x in Q20 bases; agarose-ff  
Quality coverage: 10.49x in Q20 bases; sum-of-contigs

\*\*\*\*\*  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1  
\* 64284: contig of 64284 bp in length  
\* 64285: gap of unknown length  
\* 64385: contig of 25581 bp in length  
\* 89965: gap of unknown length  
\* 90065: contig of 41904 bp in length.  
\* Location/Qualifiers  
1..131969  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="GSTNB-15B16"  
/clone\_id="GSTNB"  
/note="Genoscope designation: COAB015B16"  
1..64284  
/note="assembly-fragment  
missing Sp6 clone end on 5' end of insert"  
64385..89965  
/note="assembly-fragment"  
90065..131969  
/note="assembly-fragment  
clone\_end:T7  
vector\_side:right"

BASE COUNT 36937 a 29847 c 29144 g 35841 t 200 others  
ORIGIN

Query Match 1.2%: Score 24; DB 2; Length 131969;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2032  
Db 102320 CCAATATAGAAAAA 102297

RESULT 26  
AC010808 143655 bp DNA linear HTG 22-APR-2000  
LOCUS  
DEFINITION  
Homo sapiens clone RP11-3C7, WORKING DRAFT SEQUENCE, 3 unordered  
pieces.  
ACCESSION  
AC010808 GI:7637259  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
Homo sapiens.  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 143655)  
AUTHORS  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE  
Unpublished  
JOURNAL  
2 (bases 1 to 143655)  
REFERENCE  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgatter, B.,  
Brown, A., Castle, A., Collins, S., Collins, S., Collins, S., Collins, S.,  
Cooke, P., Dearrell, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Horton, L.,  
Galligan, J., Gaidyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howard, J., Johnson, R., Jones, C., Kahn, L., Karatas, A., Klein, J.,  
Lewoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
McEwan, P., McGurk, A., McKernan, K., McDonald, J., Meldrum, D.,  
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testaye, S., Turrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,



TITLE  
JOURNAL  
COMMENT

Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 22, 2000 this sequence version replaced gi:6454022.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L2713

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator BigDye, 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 125929 bases at least Q40

Consensus quality: 135995 bases at least Q30

Insert size: 14000; agarose-gel

Insert size: 14345; sum-of-contigs

Quality coverage: 5.2 in Q20 bases; agarose-gel

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 27688: contig of 27688 bp in length

27789 27788: gap of 100 bp

74811 74810: contig of 47022 bp in length

74911 74910: gap of 100 bp

74911 143655: contig of 68745 bp in length.

Location/Qualifiers

1. 143655

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="RP11-3C7"

/clone\_lib="RP11-11 Human Male BAC"

1. 27688

/note="assembly-fragment"

27789. 74810

/note="assembly-fragment"

clone\_end:SP6

vector\_side:left"

74911.143655

/note="assembly-fragment"

clone\_end:77

vector\_side:left"

BASE COUNT

45361 a 25207 c 25146 g 47741 t 200 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Fred. No. 0.12;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATAGCAAAAAA 2031

|||||

Db 110631 CCCAATAGCAAAAAA 110654

|||||

RESULT 27

AC073958/c

LOCUS

DEFINITION

AC073958

Homo sapiens BAC clone RP11-531L22 from 7, complete sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC073958

AC073958.4 GI:13162554

HTG.

Homo sapiens.

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 168991)

Sullivan, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

98063792

9847074

2 (bases 1 to 168991)

Goya, E., Maupin, R. and Garrett, J.

The sequence of Homo sapiens BAC clone RP11-531L22

Unpublished

3 (bases 1 to 168991)

Waterston, R.H.

Direct Submission

Submitted (07-JUN-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 168991)

Waterston, R.H.

Direct Submission

Submitted (28-FEB-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 168991)

Waterston, R.H.

Direct Submission

Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 168991)

Waterston, R.

Direct Submission

Submitted (10-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Feb 28, 2001 this sequence version replaced gi:9838250.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: sapiens@wustl.wustl.edu

----- Summary Statistics

Center project name: H\_NH0533L22

-----

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and

sequencing collaboration between the NHRI Chromosome 7 Mapping

Project (Eric D. Green, Director), John D. McPherson in the

Department of Genetics (Washington University), and the Washington

University Genome Sequencing Center. For additional information

about the map position of this sequence, see

http://www.nih.gov/DIR/STB/CHRT, send

mailto:cgreen@nhri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tano, M., Caranese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-204G8, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-533L22;

actual end is at base position 6126 of RP11-204G8.

#### FEATURES

Source

```

1..168991
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7"
  /clone="RP11-533L22"
  /clone_lib="RPCT-11"
  1464..1580
    /rpt_family="L1"
  1676..1807
    /rpt_family="Alu"
  1854..2111
    /rpt_family="Alu"
  3316..3410
    /rpt_family="MTR"
  4346..4605
    /rpt_family="L1"
  4608..5284
    /rpt_family="L1"
  5305..5377
    /rpt_family="L1"
  5367..5627
    /rpt_family="L1"
  5619..6477
    /rpt_family="L1"
  6517..8268
    /rpt_family="L1"
  8372..8459
    /rpt_family="A-rich"
  9222..9465
    /rpt_family="MaLR"
  9907..10248
    /rpt_family="ERV1"
  10292..10426
    /rpt_family="MTR"
  10462..10754
    /rpt_family="MaLR"
  11266..11330
    /rpt_family="GA-rich"
  12412..12472
    /rpt_family="MERL_type"
  12731..13304
    /rpt_family="ERV1"
  13579..14191
    /rpt_family="ERV1"
  14271..14798
    /rpt_family="ERV1"
  14826..14914
    /rpt_family="MTR"
  14939..14966
    /rpt_family="AT-rich"
  14976..15080
    /rpt_family="(TAATG)n"
  15081..15377
    /rpt_family="Alu"
  15382..15803
    /rpt_family="MERL_type"
  15891..15919
    /rpt_family="(A)n"

```

```

repeat_region 15960..16333
  /rpt_family="Alu"
repeat_region 17258..17302
  /rpt_family="(TCTA)n"
repeat_region 17322..17536
  /rpt_family="(TA)n"
repeat_region 18033..18101
  /rpt_family="L2"
repeat_region 18594..18680
  /rpt_family="(TATATG)n"
repeat_region 20418..20469
  /rpt_family="AT-rich"
repeat_region 20525..20552
  /rpt_family="AT-rich"
repeat_region 20588..20609
  /rpt_family="AT-rich"
repeat_region 21182..21440
  /rpt_family="Alu"
repeat_region 21680..21858
  /rpt_family="MERL_type"
repeat_region 22004..22317
  /rpt_family="Alu"
repeat_region 22729..22771
  /rpt_family="AT-rich"
repeat_region 22952..22974
  /rpt_family="(CAAA)n"
repeat_region 25964..25995
  /rpt_family="AT-rich"
repeat_region 27007..27028
  /rpt_family="(T)n"
repeat_region 28716..28765
  /rpt_family="AT-rich"
repeat_region 29559..29966
  /rpt_family="MaLR"
repeat_region 31352..31791
  /rpt_family="MaLR"
repeat_region 31976..32469
  /rpt_family="MaLR"
repeat_region 32690..32736
  /rpt_family="(CA)n"
repeat_region 33448..33625
  /rpt_family="L2"
repeat_region 33912..33455
  /rpt_family="L1"
repeat_region 34456..34747
  /rpt_family="Alu"
repeat_region 34748..35738
  /rpt_family="L1"
repeat_region 35712..36285
  /rpt_family="L1"
repeat_region 36350..36604
  /rpt_family="L1"

```

Query Match 1.2% Score 24; DB 9; Length 168991;  
 Best Local Similarity 100.0%; Pred. No. 0.12;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2008 CCCAATATAGAAAAA 2031  
 DB 151944 CCCAATATAGAAAAA 151921

RESULT 28  
 AC072055 AC072055 181557 bp DNA linear HTG 23-SEP-2000  
 LOCUS Homo sapiens chromosome 7 clone RP11-48C19, WORKING DRAFT SEQUENCE,  
 DEFINITION 4 unordered pieces.  
 ACCESSION AC072055  
 VERSION AC072055.3 GI:10280931  
 KEYWORDS HTG; HTGS; PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 181557)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 181557)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 On Sep 23, 2000 this sequence version replaced gi:9638044.

COMMENT ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Project Information -----  
 Center project name: H\_NH0048C19  
 Summary Statistics -----  
 Sequencing vector: M13; 73%  
 Sequencing vector: plasmid; 27%  
 Chemistry: Dye-terminator Big Dye; 73% of reads  
 Chemistry: Dye-terminator Big Dye; 27% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 180582 bases at least Q40  
 Consensus quality: 181076 bases at least Q30  
 Consensus quality: 181385 bases at least Q20  
 Insert size: 163000; agarose-fp  
 Insert size: 183250; sum-of-ctrls  
 Quality coverage: 7.87 in Q20 bases; agarose-fp  
 Quality coverage: 8.14 in Q20 bases; sum-of-ctrls

----- NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----

1 20897: contig of 20897 bp in length  
 20898 20997: gap of unknown length  
 \* 20998 50690: contig of 29693 bp in length  
 \* 50691 50790: gap of unknown length  
 \* 50791 98248: contig of 47458 bp in length  
 \* 98249 98348: gap of unknown length  
 \* 98349 181557: contig of 83209 bp in length.  
 Location/Qualifiers  
 1. 181557  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /clone="RP11-48C19"  
 1. 20897  
 /note="assembly\_name:Contig12"  
 20998 50690  
 /note="assembly\_name:Contig13"  
 50791 98248  
 /note="assembly\_name:Contig14  
 clone\_end:SP6  
 vector\_side:left"  
 98349 181557  
 /note="assembly\_name:Contig15  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 59633 a 31324 c 31034 g 59262 t 304 others  
 ORIGIN  
 Query Match 1.2%; Score 24; DB 2; Length 181557;  
 Best Local Similarity 100.0%; Pred. No. 0.12;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATACGAAAAA 2031  
 Db 114989 CCCAATACGAAAAA 115012

RESULT 29  
 AC116046  
 LOCUS  
 DEFINITION  
 Papio hamadryas clone RP41-343P17, WORKING DRAFT SEQUENCE, 4  
 AC116046 191639 bp DNA linear HTG 23-MAR-2002  
 ordered pieces.  
 AC116046  
 AC116046.1 GI:19697485  
 VERSION  
 HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Papio hamadryas  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 Cercopitheciinae; Papio.  
 1 (bases 1 to 191639)  
 Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
 Unpublished  
 2 (bases 1 to 191639)  
 Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
 Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
 Bouknight, B., Brown, A., Camarata, J., Campioano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S.,  
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,  
 Landers, T., Lechowicz, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,  
 McCarthy, M., McKernan, P., McKernan, K., Melirim, J., Menes, L.,  
 Michora, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pletre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, V., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (23-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L12531  
 Center clone name: 343\_P-17  
 Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 189809 bases at least Q40  
 Consensus quality: 190706 bases at least Q30  
 Consensus quality: 191020 bases at least Q20  
 Insert size: 192000; agarose-fp  
 Insert size: 191339; sum-of-ctrls  
 Quality coverage: 10.7 in Q20 bases; agarose-fp  
 Quality coverage: 10.7 in Q20 bases; sum-of-ctrls

----- NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 40171: contig of 40171 bp in length  
 \* 40172 40271: gap of 100 bp  
 \* 40272 70180: contig of 29909 bp in length  
 \* 70181 70280: gap of 100 bp  
 \* 70281 156363: contig of 86083 bp in length  
 \* 156364 156463: gap of 100 bp  
 \* 156464 191639: contig of 35176 bp in length.  
 Location/Qualifiers  
 1. 191639  
 /organism="Papio hamadryas"  
 /db\_xref="taxon:9557"  
 /clone="RP41-343P17"  
 /clone\_lib="RPCI-41 Male Baboon BAC"  
 1. 40171  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left"

## FEATURES

source

## misc\_feature

misc\_feature

misc\_feature

misc\_feature

BASE COUNT 55938 a 40368 c 42183 g 52844 t 306 others  
 ORIGIN

Query Match 1.2%; Score 24; DB 2; Length 191639;  
 Best Local Similarity 100.0%; Pred. No. 0.12;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCNATAGAGAAAAA 2032  
 Db 43630 CCNATAGAGAAAAA 43653

## RESULT 30

AF176841

LOCUS AF176841 1740 bp DNA linear ROD 22-SEP-1993

DEFINITION Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase

ACCESSION AF176841

VERSION AF176841.1 GI:5917711

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1740)

Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.

Cloning and characterization of a mammalian

N-acetylglucosamine-6-sulfotransferase that is highly restricted to

intestinal tissue

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

JOURNAL

MEDLINE

PUBMED

10491328

2 (bases 1 to 1740)

Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.

Direct Submission

Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,

3401 Hillview Avenue, Palo Alto, CA 94304, USA

Location/Qualifiers

1. 1740

/organism="Mus musculus"

/strain="C57Bl/6"

FEATURES

source

/db\_xref="taxon:10090"  
 /chromosome="8"  
 /map="8B1"  
 /issue\_type="intestine"  
 <405..>1592  
 /gene="I-GlcNAc-6-ST"  
 <405..>1592  
 /gene="I-GlcNAc-6-ST"  
 /product="N-acetylglucosamine 6-O-sulfotransferase"  
 405..1592  
 /gene="I-GlcNAc-6-ST"  
 /note="possibly involved in biosynthesis of mucosal  
 ligands for L-selectin"  
 /codon\_start=1  
 /product="N-acetylglucosamine 6-O-sulfotransferase"  
 /protein\_id="AAB56003.1"  
 /db\_xref="GI:5917712"

CDS

BASE COUNT 365 a 513 c 491 g 371 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 1740;  
 Best Local Similarity 100.0%; Pred. No. 0.38;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 GTGCTCAGAGAGGCGCTCTT 726  
 Db 918 GTGCTCAGAGAGGCGCTCTT 940

## RESULT 31

AF176840

LOCUS AF176840 1989 bp mRNA linear ROD 22-SEP-1993

DEFINITION Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase

ACCESSION AF176840

VERSION AF176840.1 GI:5917709

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1989)

Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.

Cloning and characterization of a mammalian

N-acetylglucosamine-6-sulfotransferase that is highly restricted to

intestinal tissue

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

JOURNAL

MEDLINE

PUBMED

10491328

2 (bases 1 to 1989)

Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.

Direct Submission

Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,

3401 Hillview Avenue, Palo Alto, CA 94304, USA

Location/Qualifiers

1. 1989

/organism="Mus musculus"

/strain="C57Bl/6"

/db\_xref="taxon:10090"

/chromosome="8"

/map="8B1"

/issue\_type="intestine"

/gene="I-GlcNAc-6-ST"

CDS

107..1294

```

/gene="I-GlcNAc-6-ST"
/notes="possibly involved in biosynthesis of mucosal
ligands for L-selection"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAB56002.1"
/db_xref="GI:5917710"
/translation="MRLPRFSVTMLSLMVOTGILVFLVSRQVPSAPAGIGRHVHL
VLSWRGSSFFVQLSQHPDYVYIMEPAHWVDI.LSQGAPALHMAVDLIHRSVPLC
DMVDFAYLPMRRNISDLFQWAVRALCPVCEAFAPGNISSEECPLCARPFG
AOCACSSSHVYLKEVRNQLQVLYPLSPALNLRVHLVRDPRAVLRROTAKAL
ARDNGYLTGNGTWEADPRLRVNEVCRRSHVRIAAALHKPPETQDRRLRYRIDL
ARPDLYIRRLIYFTGILTPLOLQIWIHNTIGSGFGARREARKTTSRDLVSQAMR
HTLPEAKIRVQELCGALQTLGYSVSHSELEFQDLSTLLPRGMDSRFMASTERQ
PES"

BASE COUNT      415 a      586 c      550 g      438 t
ORIGIN
Query Match      1.1%  Score 23;  DB 10;  Length 1989;
Best Local Similarity 100.0%;  Pred. No. 0.38;
Matches 23;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      704  GTGCTCAAGAGGTGCGCTTCTT 726
      |||||
Db      620  GTGCTCAAGAGGTGCGCTTCTT 642

RESULT 32
BC011365
LOCUS      BC011365      2002 bp      mRNA      linear      PRI 30-JUL-2001
DEFINITION Homo sapiens, clone MGC:17245 IMAGE:4182593, mRNA, complete cds.
ACCESSION  BC011365
VERSION     BC011365.1  GI:15030211
KEYWORDS   MGC.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2002)
AUTHORS   Strausberg, R.
TITLE     Direct Submission
JOURNAL   Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAX Plate: 12 Row: 1 Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1..2002
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:17245 IMAGE:4182593"
/tissue_type="Brain, glioblastoma with EGFR amplification"
/clone_lib="NCI_CGAP_Btn64"

FEATURES
source

```

```

/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
77..631
/codon_start=1
/product="Unknown (protein for MGC:17245)"
/protein_id="AAH11365.1"
/db_xref="GI:15030212"
/translation="MAEQPPSGULDEALSYCSADAPSTKDFLLQOTMLRVKDPK
SIDFTYRVLGMLTICKDPPIKESLYFLAYEDKNDIPKEDKLTAAALSRATLELT
HNMGGEDDETQSYHNGNSDPRGEGHIGIAVPYVSACRFEELGKFKVKKPPDGKMG
LAFIDPDQVTEIITNPKMALTM"

BASE COUNT      609 a      383 c      401 g      609 t
ORIGIN
Query Match      1.1%  Score 23;  DB 9;  Length 2002;
Best Local Similarity 100.0%;  Pred. No. 0.38;
Matches 23;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      2010  CAATATAGAAAAA 2032
      |||||
Db      1975  CAATATAGAAAAA 1997

RESULT 33
BC013482
LOCUS      BC013482      2459 bp      mRNA      linear      ROD 07-AUG-2002
DEFINITION Mus musculus, clone MGC:18906 IMAGE:4240772, mRNA, complete cds.
ACCESSION  BC013482
VERSION     BC013482.1  GI:15488669
KEYWORDS   MGC.
SOURCE     house mouse.
ORGANISM   Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 2459)
AUTHORS   Strausberg, R.
TITLE     Direct Submission
JOURNAL   Submitted (04-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mod@axll.stanford.edu
Dickson, M., Schultz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAX Plate: 25 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1..2459
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:18906 IMAGE:4240772"
/tissue_type="Kidney, normal, 5 month old male mouse."
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B"
/notes="Vector: PCMV-SFORP6"
42..1013
/codon_start=1
/product="Unknown (protein for MGC:18906)"

FEATURES
source
CDS

```

/db\_xref="GI:1548670"  
 /translation="MNSVSPRVVNDGHPFIPALGQTTVPDKVPKDELKATKAIADT  
 GFRHDSAYLVQIEEVEGQAIRSKIEDGTVRKEDITFTSKLWSTFHPRELVRSOLEKT  
 LKNAQLDVVDLYIHFPMALQDGRKLEPDEHGRKLAEAVDLDPTWAMEKCDAGLA  
 KSTGVSNFNPFOLETLINKPGLKYPVNCVQVCHILYNOSMDLYCKSKDILLVSYCT  
 LGSRRKTIWVQKSPVLDDPVLICAMANNKOTPALAIARYOLOGIVLITRSEKER  
 IKEMKVFEPFLASEDMKVLDDGHRNLRNTASYFDHPHPHFIDEY"

BASE COUNT 775 a 423 c 531 g 730 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 2459;  
 Best Local Similarity 100.0%; Pred. No. 0.38;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAAAAA 2032  
 Db 2436 CAAATAGAAAAA 2458

RESULT 34  
 BC026628 2488 bp mRNA linear ROD 07-AUG-2002  
 LOCUS Mus musculus, expressed sequence AW557061, clone MGC:37449  
 DEFINITION IMAGE:4983237, mRNA, complete cds.  
 ACCESSION BC026628  
 VERSION BC026628.1 GI:20071841  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 Strausberg, R.

REFERENCE  
 AUTHORS Direct Submission  
 TITLE Submitted (02-APR-2002) National Institutes of Health, Mammalian  
 JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK  
 COMMENT NIH-MGC Project URL: http://mgc.ncl.nih.gov  
 Contact: MGC help desk  
 Email: gcapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: http://www.shgc.stanford.edu  
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
 Series: IRAX Plate: 59 Row: n Column: 3  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis.

FEATURES  
 source

1. 2488  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="MGC:37449 IMAGE:4983237"  
 /tissue\_type="Colon, normal, 5 month old male mouse."  
 /clone\_lib="NCI\_CGAP\_C024"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6"  
 41. 1012  
 /codon\_start=1  
 /product="expressed sequence AW557061"  
 /protein\_id="AAH26628.1"  
 /db\_xref="GI:20071842"

BASE COUNT 805 a 422 c 531 g 730 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 2488;  
 Best Local Similarity 100.0%; Pred. No. 0.39;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAAAAA 2032  
 Db 2435 CAAATAGAAAAA 2457

RESULT 35  
 E12442/c 2860 bp DNA linear PAT 27-APR-1998  
 LOCUS DNA encoding inter site of restriction enzyme between PstI and SmaI  
 DEFINITION which contain O-acetylhomoserine sulphydrylase.  
 ACCESSION E12442  
 VERSION E12442.1 GI:3251275  
 KEYWORDS JP 1996336391-A/2.  
 SOURCE Acromonium chrysoeum.  
 ORGANISM Acromonium chrysoeum  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acromonium.  
 REFERENCE  
 AUTHORS Matsuda, A. and Muramatsu, S.  
 TITLE O-ACETYLHOMOSERINE SULPHYDRYLASE GENE  
 JOURNAL Patent: JP 1996336391-A 2 24-DEC-1996;  
 ASahi CHEM IND CO LTD  
 OS Acromonium chrysoeum  
 PN JP 1996336391-A/2  
 PD 24-DEC-1996  
 PF 13-JUN-1995 JP 1995145866  
 PI MATSUDA AKIO, MURAMATSU SHUJI  
 PC C12N15/09, C07H21/04, C12N1/15//C12N9/88, (C12N1/15, C12R1:645),  
 PC (C12N9/88,  
 PC C12R1:645);  
 CC strandedness: Double;  
 CC topology: linear;  
 CC hypothetical: No;  
 CC anti-sense: No;  
 FH key location/Qualifiers  
 EH source 1. 2860  
 FT /organism="Acromonium chrysoeum" FT  
 FT /strain="IS-5"  
 FT 5'UTR 1. 448  
 FT mat-peptide 449. 495  
 FT /product="inter site of restriction enzyme  
 FT between PstI and  
 FT SmaI which contain O-acetylhomoserine FT  
 FT sulphydrylase"  
 FT mat-peptide 647. 701  
 FT /product="inter site of restriction enzyme  
 FT between PstI and  
 FT SmaI which contain O-acetylhomoserine FT  
 FT sulphydrylase"  
 FT mat-peptide 1316. 1914  
 FT /product="inter site of restriction enzyme  
 FT between PstI and  
 FT SmaI which contain O-acetylhomoserine FT

```

FEATURES
  source
    1. .2860
    /organism="Acremonium chrysogenum"
    /db_xref="taxon:5044"
BASE COUNT      570 a      933 c      766 g      591 t
ORIGIN
Query Match      1.1%; Score 23; DB 6; Length 2860;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032
|||||
Db 1248 CAATAGAGAAAAA 1226

RESULT 36
AC017237/c
LOCUS      AC017237      29064 bp      DNA      linear      HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION  AC017237
VERSION    AC017237.1 GI:6553749
KEYWORDS   HTG; HTGS; PHASE2.
SOURCE     Drosophila melanogaster.
ORGANISM   Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 29064)
            Adams, M. and Venter, J.C.
            Direct Submission
            Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
            This sequence was identified as CDM:10209972 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES
  source
    1. .29064
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
BASE COUNT      8648 a      5944 c      5998 g      8474 t
ORIGIN
Query Match      1.1%; Score 23; DB 2; Length 29064;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

RESULT 37
AC116989
LOCUS      AC116989      55470 bp      DNA      linear      HTG 24-MAY-2002
DEFINITION Dictyostelium discoideum chromosome 2 map 3542416-3597884 strain
AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION  AC116989
VERSION    AC116989
KEYWORDS   HTG; HTGS; PHASE2.
SOURCE     Dictyostelium discoideum.
ORGANISM   Dictyostelium discoideum.
            Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
REFERENCE  1 (bases 1 to 55470)
            Gloeckner, G., Eichinger, L., Szafirski, K., Pachebat, J., Dear, P.,
            Lehmann, R., Baumgart, C., Parra, G., April, J.F., Gulgo, R., Kumpf, K.,
            Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A. and
            Noegel, A.A.
            Sequence and Analysis of Chromosome 2 of Dictyostelium
            Unpublished
            The Dictyostelium Genome Sequencing Consortium
            2 (bases 1 to 55470)
            Baumgart, C.
            Direct Submission
            Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
            3 (bases 1 to 55470)
            Baumgart, C.
            Direct Submission
            Submitted (24-MAY-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
            CDS predictions from GeneID may contain errors. Further information
            is available from IMB Jena, Department of Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the University Cologne, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml)
            Funding
            Agency: Deutsche Forschungsgemeinschaft (DFG).
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES
  source
    1. .55470
    /organism="Dictyostelium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    /chromosome="2"
    /map="3542416-3597884"
    2458. .6183
    /note="ORF_ID: d4_03152"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AAM43736.1"
    /db_xref="GI:21166119"
    /translation="MEPDEETRIEMERRIRPERRKQEDDAKEDORRKAEREKRY
    KREATKREGEREAQKKLEBYTRTEERRRQREERREKREERKQKDEERK
    QOERERKREDEERKRIKLEESRKVLELEDENSSQVDSISQKRRRLIARQ
    KTESILAEVDPSPNSQYDQTPPIKIVSLLITDELKEVYDILVAGGSRVDPVSO
    ASISTMAEIVDINVGATIFASKPEFLVGTASLEKISLARMOGAKSPILLIYCNKQITL
    ITDPSNFIIDILFMSINPPIILNMHIETDLDVCSYMSRKPILLIYCNKQITL
    PIMQHSNSEVEDVLLKMAETEMQSNSETSSGSAASFESEITSGSSSSSS
    NNNNNNNKKERPNMTNDLTLYILIRINISLEAVYDIDIPSTIARLMFELYK
    YKSVVAHLICGQSTSNVIRNGEFLALVHPTSPFSMSLITLIVDSVYTMQSL
    ITNSSSSSSSLLSSSSSSSLLSSSSSSNDLPVNLKQIVITIDISNGEDNSR
    ELSPKALSEIITIKRDSQELFIVRLCEHLINSLVANKSKIDREIWSGIKIKI
    DLFEPYETNNVYHCHMIGILTPIIORLGSDEDEPMCLLDCSFVSVVILSGNN
    SELNDINNSNNNNNNNNNNNNNNNNNNNGSLGASKSLATFVSGKNDVYIE
    OOKSEPVIGAPAPAMGSLSRGGKKIRKRVSEFSDSTSNADQENNTSNTNGDVK
    IALGIDLMQPTSTPAPATSTTBOSSNGLNENYVSSERKFAITLSTSISSSS
    SSTSTSTSTTTTKINGTSSSSSSSSSSSKAKAKELKQDEKRIKELMDTKSKO
    LTPDQEKAKSQIKQLKREKREKREKRLAKRLKREKSKSTSTSTSLKNS
    GSVVLPSTINKLKSSSSNSPSSSPVLTSTTSLAAATTTTTPSTTAPPTTTT
    TMTTITITSLSDQSDTSNSDSCNSDVSEIEIKLINNSQPIEASVNTISIEE
    OSVVVVYHNEESSNNGSSNTECNNSKINDESGTNTPTTSTTTAKKGVHNGHII
```

CDS		complement(join(16864..17173,17523..17530)) /note="ORF_ID:dq_01367" /codon_start=1 /product="hypothetical protein" /protein_id="AAM43744.1" /db_xref="GI:21166127" /translation="MLEPFLKKNFINVLEFNKKLPSPSTIGNVKSISNNFSSLSN /translationalion="MLEPFLKKNFINVLEFNKKLPSPSTIGNVKSISNNFSSLSN SLSLGSNSISIQCGGGGSGPLISTGVNLVGVLWGNIIITIGVTGTVNGVGLLSGPNC GCH"
CDS		join(17865..17877,17959..18215) /note="ORF_ID:dq_01388" /codon_start=1 /product="hypothetical protein" /protein_id="AAM43745.1" /db_xref="GI:21166128" /translation="MTILASISSIGNVKSIKSNPFSSLNSLSLOSSNSTIOCGCGGGG NSLIGTAVGLVGVAGLVGTGIVTGTVGVNGVNGVGLLSGPNCGCH" complement(join(18633..18888,18961..18973)) /note="ORF_ID:dq_01385" /codon_start=1 /product="hypothetical protein" /protein_id="AAM43746.1" /db_xref="GI:21166129" /translation="MTILASISSIGNVKSIKSNPFSSLNSLSLOSSNSTIOCGCGGGG SPLIGTAVGLVGVAGLVGTGIVTGTVGVNGVNGVGLLSGPNCGCH" join(19666..19678,19759..20015) /note="ORF_ID:dq_01382" /codon_start=1 /product="hypothetical protein" /protein_id="AAM43747.1" /db_xref="GI:21166130" /translation="MTILASISSIGNVKSIKSNPFSSLNSLSLOSSNSTIOCGCGGGG SPLIGTAVGLVGVAGLVGTGIVTGTVGVNGVNGVGLLSGPNCGCH" complement(join(20170..20736,20806..20864,20943..21345)) /note="ORF_ID:dq_01379" /codon_start=1 /product="putative oxidoreductase" /protein_id="AAM43748.1" /db_xref="GI:21166131" /translation="MGVFPTDISLTITNIGSIILLIPIYLPIQNKYKKKKTNPTPA LIVTGASSGICRGIALLEYAKRRKNNSVLGITGRNIKELETOKECIGLYOVEIESI DVTDEPKLNDWLIKFDNKRYKIDILIANAGTIETLLPELDTERILTVAINTVATGIN TVLPEVPFIRESGGQALWMSITPYEDYVAAGYSSSKGYIKSGELLRGLARVGVS VSIVIPGRIPTPLVDSDLYKDIPMLPDTTKASKAYVDGISRDALIISPPLIYCYS HEVATIPETPLDATNFINSLSLKYPDKSPVTHXSHHSOYSTNSKRREDCPSAANMS NSOKRKA"
CDS		21759..22802 /note="ORF_ID:dq_01376" /codon_start=1 /product="cysteine-rich, acidic integral membrane protein precursor" /protein_id="AAM43749.1"
Query Match	1.1%	Score 23; DB 2; Length 55470; Best local similarity 100.0%; Pred. No. 0.45;
Matches	23; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	2010 CAAATACAGAAAAA	2032
Db	53222 CAATACAGAAAAA	53244
RESULT 38	AC103270	56342 bp DNA linear HTG 13-JUL-2002
LOCUS	AC103270/c	Rattus norvegicus clone CH230-228B1, *** SEQUENCING IN PROGRESS
DEFINITION	*** 35 unrounded pieces.	
ACCESION	AC103270.3	GI:21731258



## KEYWORDS

HTG: HTGS PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

## REFERENCE

1 (bases 1 to 56342)

Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralting,H.C., Are,J.R., Ayale,M., Banks,T.,

Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonin,D.,

Burch,J., Bowler,S., Brivela,M., Brown,E., Brown,M., Bryant,N.P.,

Butay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Daborne,S.R., David,R.,

DeLaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

DeLaney,R.R., Delgado,O., Denn,A.L., Ding,Y., Dink,H.H.,

Doutlaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogue,M., Hollway,C., Hollins,B.,

Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Licharge,O., Lieu,C., Liu,T., Liu,W., Lounseged,H.,

Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Meheswari,M., Mapua,P., Martin,R., Matlindale,A., Martinez,E.,

Massey,E., Mawney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,

Minner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,S., Nguy,N., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okunon,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Peter,L.,

Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,

Rever,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G.,

Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,

Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Uemai,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R., Wang,Q.,

Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,

Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Unpublished

Direct Submission

2 (bases 1 to 56342)

Worley,K.C.

Direct Submission

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 56342)

Worley,K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 11, 2002 this sequence version replaced gi:17974782.

\*\*\*\*\* Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

\*\*\*\*\* Project Information

Center project name: GJ02

Center clone name: CH230-228B1

\*\*\*\*\* Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329  
Consensus quality: 9543 bases at least Q40  
Consensus quality: 10073 bases at least Q30  
Consensus quality: 10542 bases at least Q20  
\*\*\*\*\*

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 35 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1	1205:	contig of 1205 bp in length
1206	1305:	gap of unknown length
1306	2373:	contig of 1068 bp in length
2374	2473:	gap of unknown length
2474	3564:	contig of 1091 bp in length
3565	3664:	gap of unknown length
3665	4730:	contig of 1066 bp in length
4731	4830:	gap of unknown length
4831	6009:	contig of 1179 bp in length
6010	6109:	gap of unknown length
6110	7354:	contig of 1245 bp in length
7355	7454:	gap of unknown length
7455	8697:	contig of 1243 bp in length
8698	8797:	gap of unknown length
8798	9797:	contig of 1000 bp in length
9798	9897:	gap of unknown length
9898	10909:	contig of 1012 bp in length
10910	11009:	gap of unknown length
11010	12689:	contig of 1680 bp in length
12690	12789:	gap of unknown length
12790	14479:	contig of 1630 bp in length
14480	14579:	gap of unknown length
14580	15724:	contig of 1145 bp in length
15725	15824:	gap of unknown length
15825	16946:	contig of 1122 bp in length
16947	17046:	gap of unknown length
17047	18079:	contig of 1033 bp in length
18080	18179:	gap of unknown length
18180	19855:	contig of 1676 bp in length
19856	19955:	gap of unknown length
19956	20982:	contig of 1027 bp in length
20983	21082:	gap of unknown length
21083	22928:	contig of 1846 bp in length
22929	23028:	gap of unknown length
23029	24033:	contig of 1005 bp in length
24034	24133:	gap of unknown length
24134	25243:	contig of 1110 bp in length
25244	25343:	gap of unknown length
25344	27165:	contig of 1822 bp in length
27166	27265:	gap of unknown length
27266	28838:	contig of 1573 bp in length
28839	28938:	gap of unknown length
28939	30372:	contig of 1434 bp in length
30373	30472:	gap of unknown length
30473	32069:	contig of 1597 bp in length
32070	32169:	gap of unknown length
32170	34053:	contig of 1884 bp in length
34054	34153:	gap of unknown length
34154	35815:	contig of 1662 bp in length
35816	35915:	gap of unknown length
35916	37432:	contig of 1517 bp in length
37433	37532:	gap of unknown length
37533	39072:	contig of 1540 bp in length
39073	39172:	gap of unknown length
39173	40536:	contig of 1367 bp in length
40536	40639:	gap of unknown length
40640	42394:	contig of 1755 bp in length
42395	42494:	gap of unknown length

```

* 42495 44274: contig of 1780 bp in length
* 44275 44374: gap of unknown length
* 44375 46372: contig of 2198 bp in length
* 46373 46673: gap of unknown length
* 46673 47937: contig of 1265 bp in length
* 47938 48037: gap of unknown length
* 48038 49789: contig of 1762 bp in length
* 49800 49899: gap of unknown length
* 49900 51038: contig of 1139 bp in length
* 51039 51138: gap of unknown length
* 51139 56342: contig of 5204 bp in length.
  Location/Qualifiers
    source          .56342
    organism        "Rattus norvegicus"
    db_xref="taxon:10116"
    /clone="CH230-228B1"

BASE COUNT      15092 a 11063 c 11162 g 13872 t 5153 others
ORIGIN

Query Match      1.18; Score 23; DB 2; Length 56342;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032
|||||
Db 6829 CAATAGAGAAAAA 6807

RESULT 39
AC087621 63739 bp DNA linear HTG 15-JAN-2001
LOCUS      Homo sapiens chromosome 8 clone RP11-300E4 map 8, LOW-PASS SEQUENCE
DEFINITION
AC087621
AC087621.1 GI:12229396
HTG: HTGS_PHASE0.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 63739)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 8, clone RP11-300E4
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 63739)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Camataia,J., Campopiano,A., Choquet,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glade,S., Goyette,M., Graham,L., Grenier-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Karas,A., Larocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Margolis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,C., Norbu,C., Norman,C.H.,
Nurphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Plunkhant,P., Pierre,N., Pollara,V., Raymond,C., Retten,R.,
Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,
Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Sudamanian,A., Talamas,J., Testaie,S., Theodore,J.,
Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zambek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/rw/RepeatMasker.html
----- Genome Center

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11683
Center clone name: 300_E4
-----
* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 701 800: gap of 100 bp
* 801 1491: contig of 691 bp in length
* 1492 1591: gap of 100 bp
* 1592 2304: contig of 713 bp in length
* 2305 2404: gap of 100 bp
* 2405 3109: contig of 705 bp in length
* 3110 3209: gap of 100 bp
* 3210 3926: contig of 717 bp in length
* 3927 4026: gap of 100 bp
* 4027 4745: contig of 719 bp in length
* 4746 4845: gap of 100 bp
* 4846 5573: contig of 728 bp in length
* 5574 5673: gap of 100 bp
* 5674 6395: contig of 722 bp in length
* 6396 6495: gap of 100 bp
* 6496 7204: contig of 709 bp in length
* 7205 7304: gap of 100 bp
* 7305 8014: contig of 710 bp in length
* 8015 8114: gap of 100 bp
* 8115 8828: contig of 714 bp in length
* 8829 8928: gap of 100 bp
* 8929 9622: contig of 694 bp in length
* 9623 9722: gap of 100 bp
* 9723 10431: contig of 709 bp in length
* 10432 10531: gap of 100 bp
* 10532 11330: contig of 699 bp in length
* 11331 11330: gap of 100 bp
* 11331 11870: contig of 640 bp in length
* 11971 12070: gap of 100 bp
* 12071 12823: contig of 753 bp in length
* 12824 12923: gap of 100 bp
* 12924 13630: contig of 707 bp in length
* 13631 13730: gap of 100 bp
* 13731 14448: contig of 718 bp in length
* 14449 14548: gap of 100 bp
* 14549 15262: contig of 714 bp in length
* 15263 15362: gap of 100 bp
* 15363 16076: contig of 714 bp in length
* 16077 16176: gap of 100 bp
* 16177 16880: contig of 704 bp in length
* 16881 16980: gap of 100 bp
* 16981 17717: contig of 737 bp in length
* 17718 17817: gap of 100 bp
* 17818 18522: contig of 705 bp in length
* 18523 18622: gap of 100 bp
* 18623 19343: contig of 721 bp in length
* 19344 19443: gap of 100 bp
* 19444 20161: contig of 718 bp in length
* 20162 20261: gap of 100 bp
* 20262 20983: contig of 722 bp in length
* 20984 21083: gap of 100 bp
* 21084 21810: contig of 727 bp in length
* 21811 21910: gap of 100 bp

```

21911 22637: contig of 727 bp in length  
22638 22737: gap of 100 bp  
22738 23439: contig of 702 bp in length  
23440 23539: gap of 100 bp  
23540 24247: contig of 708 bp in length  
24248 24347: gap of 100 bp  
24348 25051: contig of 704 bp in length  
25052 25151: gap of 100 bp  
25152 25851: contig of 700 bp in length  
25852 25951: gap of 100 bp  
25952 26675: contig of 724 bp in length  
26676 26775: gap of 100 bp  
26776 27470: contig of 695 bp in length  
27471 27570: gap of 100 bp  
27571 28264: contig of 694 bp in length  
28265 28364: gap of 100 bp  
28365 29085: contig of 721 bp in length  
29086 29185: gap of 100 bp  
29186 29892: contig of 707 bp in length  
29893 29993: gap of 100 bp  
29993 30703: contig of 711 bp in length  
30704 30803: gap of 100 bp  
30804 31509: contig of 706 bp in length  
31510 31609: gap of 100 bp  
31610 32297: contig of 688 bp in length  
32298 32397: gap of 100 bp  
32398 33104: contig of 707 bp in length  
33105 33204: gap of 100 bp  
33205 33905: contig of 701 bp in length  
33906 34005: gap of 100 bp  
34006 34708: contig of 703 bp in length  
34709 34808: gap of 100 bp  
34809 35513: contig of 705 bp in length  
35514 35613: gap of 100 bp  
35614 36345: contig of 732 bp in length  
36346 36445: gap of 100 bp  
36446 37174: contig of 729 bp in length  
37175 37274: gap of 100 bp  
37275 37987: contig of 713 bp in length  
37988 38087: gap of 100 bp  
38088 38800: contig of 713 bp in length  
38801 38900: gap of 100 bp  
38901 39598: contig of 698 bp in length  
39599 39698: gap of 100 bp  
39699 40403: contig of 705 bp in length  
40404 40503: gap of 100 bp  
40504 41221: contig of 718 bp in length  
41222 41321: gap of 100 bp  
41322 42007: contig of 686 bp in length  
42008 42107: gap of 100 bp  
42108 42794: contig of 687 bp in length  
42795 42894: gap of 100 bp  
42895 43612: contig of 718 bp in length  
43613 43712: gap of 100 bp  
43713 44428: contig of 717 bp in length  
44429 44529: gap of 100 bp  
44530 45239: contig of 710 bp in length  
45240 45339: gap of 100 bp  
45340 46049: contig of 710 bp in length  
46050 46149: gap of 100 bp  
46150 46864: contig of 715 bp in length  
46865 46964: gap of 100 bp  
46965 47674: contig of 710 bp in length  
47675 47774: gap of 100 bp  
47775 48465: contig of 691 bp in length  
48466 48565: gap of 100 bp  
48566 49271: contig of 706 bp in length  
49272 49371: gap of 100 bp  
49372 50018: contig of 647 bp in length  
50019 50118: gap of 100 bp  
50119 50839: contig of 721 bp in length  
50840 50939: gap of 100 bp  
50940 51664: contig of 725 bp in length

## Query Match

Best Local Similarity 1.1%; Score 23; DB 2; Length 63739;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032

Db 33706 CAATATAGAAAAA 33728

## RESULT 40

## AC124305/c

## LOCUS

## DEFINITION

## SEQUENCE

## SAMPLING

## AC124305

## AC124305.1

## VERSION

## KEYWORDS

## HTG: HTGS-PHASE0.

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## 2 (bases 1 to 69523)

## Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

## Bouckhaert, B., Brown, A., Camarata, J., Campione, A., Chang, J.,

## Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,

## Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., Dodge, S.,

## Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,

## Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L.,

## Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,

## Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K.,

## Lamas, R., Lander, T., Lehotzky, J., Levine, R., Lindblad-Toh, K.,

## Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N.,

## Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Melitini, J.,

## Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,

## Nicoli, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,

## O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

## Pollara, V., Raymond, C., Retra, R., Rieback, M., Riley, R., Rise, C.,

## Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,

## Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,

## Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,

## Theodore, J., Topham, K., Travers, M., Travis, N., Trifillio, J.,

## Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,

## Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## COMMENT

51665 51764: gap of 100 bp  
51765 52477: contig of 713 bp in length  
52478 52577: gap of 100 bp  
52578 53281: contig of 704 bp in length  
53282 53381: gap of 100 bp  
53382 54081: contig of 700 bp in length  
54082 54181: gap of 100 bp  
54182 54877: contig of 696 bp in length  
54878 54977: gap of 100 bp  
54978 55682: contig of 705 bp in length  
55683 55782: gap of 100 bp

AC124305 69523 bp DNA linear HTG 14-JUN-2002  
Homo sapiens chromosome 15 clone RP11-173H16 map 15, LOW-PASS  
SEQUENCE SAMPLING.  
AC124305.1 GI:21426329  
HTG: HTGS-PHASE0.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 69523)  
Homo sapiens chromosome 15, clone RP11-173H16  
Unpublished  
2 (bases 1 to 69523)  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
Bouckhaert, B., Brown, A., Camarata, J., Campione, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,  
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L.,  
Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K.,  
Lamas, R., Lander, T., Lehotzky, J., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N.,  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Melitini, J.,  
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicoli, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retra, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trifillio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L27334

Center Clone name: 173\_H\_16

-----  
\* NOTE: This record contains 84 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\*  
\* 1  
\* 708 807: contig of 707 bp in length  
\* 808 1544: contig of 737 bp in length  
\* 1545 1644: gap of 100 bp  
\* 1645 2355: contig of 711 bp in length  
\* 2356 2455: gap of 100 bp  
\* 2456 3198: contig of 744 bp in length  
\* 3200 3299: gap of 100 bp  
\* 3300 4006: contig of 707 bp in length  
\* 4007 4106: gap of 100 bp  
\* 4107 4838: contig of 732 bp in length  
\* 4839 4938: gap of 100 bp  
\* 4939 5676: contig of 738 bp in length  
\* 5677 5776: gap of 100 bp  
\* 5777 6490: contig of 714 bp in length  
\* 6491 6590: gap of 100 bp  
\* 6591 7307: contig of 717 bp in length  
\* 7308 7407: gap of 100 bp  
\* 7408 8136: contig of 729 bp in length  
\* 8137 8236: gap of 100 bp  
\* 8237 8975: contig of 739 bp in length  
\* 8976 9075: gap of 100 bp  
\* 9076 9817: contig of 742 bp in length  
\* 9818 9917: gap of 100 bp  
\* 9918 10665: contig of 748 bp in length  
\* 10666 10765: gap of 100 bp  
\* 10766 11510: contig of 745 bp in length  
\* 11511 11610: gap of 100 bp  
\* 11611 12353: contig of 743 bp in length  
\* 12354 12453: gap of 100 bp  
\* 12454 13195: contig of 742 bp in length  
\* 13196 13295: gap of 100 bp  
\* 13296 14021: contig of 726 bp in length  
\* 14022 14121: gap of 100 bp  
\* 14122 14839: contig of 718 bp in length  
\* 14840 14939: gap of 100 bp  
\* 14940 15646: contig of 707 bp in length  
\* 15647 15746: gap of 100 bp  
\* 15747 16482: contig of 736 bp in length  
\* 16483 16582: gap of 100 bp  
\* 16583 17280: contig of 698 bp in length  
\* 17281 17380: gap of 100 bp  
\* 17381 18109: contig of 729 bp in length  
\* 18110 18209: gap of 100 bp  
\* 18210 18949: contig of 740 bp in length  
\* 18950 19049: gap of 100 bp  
\* 19050 19783: contig of 734 bp in length  
\* 19784 19883: gap of 100 bp  
\* 19884 20602: contig of 719 bp in length  
\* 20603 20702: gap of 100 bp  
\* 20703 21439: contig of 737 bp in length  
\* 21440 21539: gap of 100 bp  
\* 21540 22271: contig of 732 bp in length  
\* 22272 22371: gap of 100 bp  
\* 22372 23075: contig of 704 bp in length  
\* 23076 23175: gap of 100 bp  
\* 23176 23875: contig of 700 bp in length  
\* 23876 23975: gap of 100 bp  
\* 23976 24708: contig of 733 bp in length  
\* 24709 24808: gap of 100 bp

\* 24809 25543: contig of 735 bp in length  
\* 25444 25643: gap of 100 bp  
\* 25644 26383: contig of 740 bp in length  
\* 26384 26483: gap of 100 bp  
\* 26484 27219: contig of 736 bp in length  
\* 27220 27319: gap of 100 bp  
\* 27320 28060: contig of 741 bp in length  
\* 28061 28160: gap of 100 bp  
\* 28161 28896: contig of 736 bp in length  
\* 28897 28996: gap of 100 bp  
\* 28997 29734: contig of 738 bp in length  
\* 29735 29834: gap of 100 bp  
\* 29835 30568: contig of 734 bp in length  
\* 30569 30668: gap of 100 bp  
\* 30669 31405: contig of 737 bp in length  
\* 31406 31505: gap of 100 bp  
\* 31506 32230: contig of 725 bp in length  
\* 32231 32330: gap of 100 bp  
\* 32331 33057: contig of 727 bp in length  
\* 33058 33157: gap of 100 bp  
\* 33158 33892: contig of 735 bp in length  
\* 33893 33992: gap of 100 bp  
\* 33993 34726: contig of 734 bp in length  
\* 34727 34826: gap of 100 bp  
\* 34827 35528: contig of 702 bp in length  
\* 35529 35628: gap of 100 bp  
\* 35629 36361: contig of 733 bp in length  
\* 36362 36461: gap of 100 bp  
\* 36462 37200: contig of 739 bp in length  
\* 37201 37300: gap of 100 bp  
\* 37301 38028: contig of 728 bp in length  
\* 38029 38128: gap of 100 bp  
\* 38129 38865: contig of 737 bp in length  
\* 38866 38965: gap of 100 bp  
\* 38966 39705: contig of 740 bp in length  
\* 39706 39805: gap of 100 bp  
\* 39806 40538: contig of 733 bp in length  
\* 40539 40638: gap of 100 bp  
\* 40639 41354: contig of 716 bp in length  
\* 41355 41454: gap of 100 bp  
\* 41455 42181: contig of 727 bp in length  
\* 42182 42281: gap of 100 bp  
\* 42282 42999: contig of 718 bp in length  
\* 43000 43099: gap of 100 bp  
\* 43100 43833: contig of 734 bp in length  
\* 43834 43933: gap of 100 bp  
\* 43934 44667: contig of 734 bp in length  
\* 44668 44767: gap of 100 bp  
\* 44768 45472: contig of 705 bp in length  
\* 45473 45572: gap of 100 bp  
\* 45573 46309: contig of 737 bp in length  
\* 46310 46409: gap of 100 bp  
\* 46410 47110: contig of 701 bp in length  
\* 47111 47210: gap of 100 bp  
\* 47211 47918: contig of 708 bp in length  
\* 47919 48018: gap of 100 bp  
\* 48019 48746: contig of 728 bp in length  
\* 48747 48846: gap of 100 bp  
\* 48847 49582: contig of 736 bp in length  
\* 49583 49682: gap of 100 bp  
\* 49683 50413: contig of 737 bp in length  
\* 50420 50519: gap of 100 bp  
\* 50520 51224: contig of 705 bp in length  
\* 51225 51324: gap of 100 bp  
\* 51325 52061: contig of 737 bp in length  
\* 52062 52161: gap of 100 bp  
\* 52162 52873: contig of 712 bp in length  
\* 52874 52973: gap of 100 bp  
\* 52974 53688: contig of 715 bp in length  
\* 53689 53788: gap of 100 bp  
\* 53789 54521: contig of 733 bp in length  
\* 54522 54621: gap of 100 bp  
\* 54622 55327: contig of 706 bp in length

\* 55328 55427: gap of 100 bp  
 \* 55428 56157: contig of 730 bp in length  
 \* 56158 56257: gap of 100 bp  
 \* 56258 56999: contig of 742 bp in length

Query Match 1.1%; Score 23; DB 2; Length 69523;  
 Best Local Similarity 100.0%; Pred. No. 0.46; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGCAAAAAAAAAAAAAA 2032  
 Db 40400 CAATTAAGCAAAAAAAAAAAAAA 40378

RESULT 41  
 AC083858 74718 bp DNA linear ROD 05-OCT-2001  
 LOCUS Mus musculus chromosome 5 clone RP23-423A22 strain C57BL6/J,  
 DEFINITION complete sequence.  
 AC083858  
 VERSION AC083858.3 GI:15963650  
 KEYWORDS HTG.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

1 (bases 1 to 74718)  
 Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Hin, S.-Q.,  
 Legaspi, R., Maduro, Q.L., Maduro, V.B., Mastaglio, C., Mastrian, S.D.,  
 McLooney, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,  
 Stantropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,  
 Tsurgren, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Zhang, L., H. and Green, E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 74718)  
 Submitted (04-OCT-2000) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 3 (bases 1 to 74718)  
 Green, E.D.  
 Direct Submission  
 Submitted (05-OCT-2001) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 On Oct 5, 2001 this sequence version replaced gi:12313754.  
 COMMENT  
 Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: http://www.nisc.nih.gov  
 Contact: nisc.mouse@nhgri.nih.gov  
 Project Information  
 Center project name: r9  
 Center clone name: 423A22

This sequence was finished as follows unless otherwise noted:  
 all regions were double-stranded, sequenced with an  
 alternate chemistry, or covered by high quality data  
 (i.e., phred quality >= 30); an attempt was made to resolve  
 all sequencing problems, such as compressions and repeats;  
 all regions were covered by at least one plasmid subclone  
 or more than one M13 subclone; and the assembly was confirmed  
 by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of  
 this clone unless otherwise noted. If there are overlapping  
 clones, the overlaps are noted in the beginning and end of  
 the features section.

FEATURES  
 Location/Qualifiers  
 1..74718

## misc-feature

/organism="Mus musculus"  
 /strain="C57BL6/J"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /clone="RP23-423A22"  
 /clone\_lib="RP23 mouse BAC library 23"  
 <1..9380  
 /note="this sequence is not the entire insert of clone  
 RP23-423A22; clone overlaps with GenBank Accession Number  
 AC024608 (nucleotides 104851-215829) clone RP23-333124  
 (center project name xy); this annotated segment  
 represents overlap with nucleotides 206450-215829 of  
 AC024608"

## misc-feature

37106..37179  
 /note="single clone coverage"  
 70746..>74718  
 /note="this sequence is not the entire insert of clone  
 RP23-423A22; clone overlaps with GenBank Accession Number  
 AC083948 (nucleotides 1-45970) clone RP23-67P7 (center  
 project name rr); this annotated segment represents  
 overlap with nucleotides 1-3973 of AC083948"

## misc-feature

BASE COUNT 19836 a 17184 c 17496 g 20202 t

## ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 0.46; Length 74718;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGCAAAAAAAAAAAAAA 2032  
 Db 40963 CAATTAAGCAAAAAAAAAAAAAA 40985

## RESULT 42

## LOCUS

AC117427/c 83543 bp DNA linear PRI 25-MAY-2002  
 DEFINITION Homo sapiens 3 BAC RP11-6615 (Roswell Park Cancer Institute Human  
 BAC Library) complete sequence.  
 AC117427  
 VERSION AC117427.2 GI:21206085  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 83543)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., All-roman, F.R., Allen, C.,  
 Alstbrooks, S.L., Anarlungue, H.C., Are, J.R., Ayele, M., Banks, T.,  
 Barbieri, J., Benton, O., Blmage, K., Blankensbury, K., Bonnin, D.,  
 Bouck, J., Bowls, S., Briley, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buha, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,  
 Carion, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dedetich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabriel, R., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorielli, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., He, X.,  
 Hernandez, J., Hernandez, O., Hoogson, A., Hogues, M., Hollway, C.,  
 Hollins, B., Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J.,  
 Kovar, C., Kralovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,  
 Louisseg, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
 Mel, G., Metker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Moore, S., Morgan, M., Morris, T., Morris, S., Mosser, M., Neal, D.,



```

repeat_region 26895..27733
/rpt_family="L1PA13"
repeat_region 28372..29543
/rpt_family="L1PA16"
repeat_region 29544..30552
/rpt_family="MER11A"
repeat_region 30553..30863
/rpt_family="L1PA16"
repeat_region 30864..30906
/rpt_family="TAA)n"
repeat_region 30907..32672
/rpt_family="L1PA16"
repeat_region complement(32673..33726)
/rpt_family="MER11C"
repeat_region 33727..34480
/rpt_family="L1PA16"

```

Query Match 1.1%; Score 23; DB 9; Length 83543;

Best Local Similarity 100.0%; Pred.No. 0.46; Mismatches 23; Conservative 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAAAAAAAAAA 2032  
 DB 73623 CAATATAGAAAAAAAAAAAAA 73601

RESULT 43  
 AL590368/c 92910 bp DNA linear PRI 15-NOV-2001  
 LOCUS Human DNA sequence from clone RP11-3K19 on chromosome 9, complete  
 DEFINITION  
 ACCESSION AL590368 AC044833  
 VERSION AL590368.11 GI:16973067  
 KEYWORDS HNG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 92910)  
 AUTHORS Bates,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Nov 16, 2001 this sequence version replaced gi:16416219.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

COMMENT  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C-elegans/wormPep This sequence  
 was generated from part of bacterial clone configs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr9  
 RP11-3K19 is from the library RPCI-11.1 constructed by the group of  
 Pieter de Jong. For further details see  
 http://www.chori.org/dacpac/home.htm  
 VECTOR: pBAC3.6  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP11-3K19 it may be shorter because we sequence overlapping

sections only once, except for a short overlap.  
 The true left end of clone RP11-477A7 is at 90911 in this sequence.  
 The true right end of clone RP11-112A13 is at 2000 in this  
 sequence.

FEATURES  
 source  
 1..92910  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-3K19"  
 /clone\_1bp="RPCI-11.1"

misc-feature  
 22581..22800  
 /note="Single clone region. Sequence from reads from a  
 short insert library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."

BASE COUNT 25801 a 18381 c 18458 g 30270 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 92910;

Best Local Similarity 100.0%; Pred.No. 0.47; Mismatches 23; Conservative 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAAAAAAAAAA 2032  
 DB 24900 CAATATAGAAAAAAAAAAAAA 24878

RESULT 44  
 HS874C20/c 97847 bp DNA linear PRI 24-JAN-2002  
 LOCUS Human DNA sequence from clone RP5-874C20 on chromosome 6p22.1-22.3.  
 DEFINITION  
 ACCESSION HS874C20  
 VERSION AL021997.1 GI:3169112  
 HNG: Cpg island; SRE-ZBP; ZFP47; zinc finger protein; ZNF306.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 97847)  
 AUTHORS Williams,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On May 30, 1998 this sequence version replaced gi:2916865.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

COMMENT  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C-elegans/wormPep This sequence  
 was generated from part of bacterial clone configs of human  
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6  
 RP5-874C20 is from the library RPCI-5 constructed by the group of  
 Pieter de Jong. For further details see

```
http://www.chori.org/bacpac/home.htm
VECTOR: pCIPAC2
This sequence is the entire insert of clone RP5-874C20.
FEATURES
    source
        1..97847
            Location/Qualifiers
                organism="Homo sapiens"
                db_xref="R2PD:RPCIP704C20874"
                db_xref="taxon:9606"
                chromosome="6"
                map="p22.1-22.3"
                clone="RP5-874C20"
                clone_lib="RPCI-5"
                1..111
repeat_region
    /note="Aluub repeat: matches 194..304 of consensus"
gene
    complement(262..4744)
    /gene="ba245E14.1"
    /complement(join(<262..385,692..774,2606..2753,4089..4744))
mRNA
    /gene="ba245E14.1"
    /product="d3874C20.1 (novel protein similar to zinc finger protein 306 (ZNF306, ZFP47))"
    /evidence=not_experimental
    /complement(join(<262..385,692..774,2606..2753,4089..4490))
    /gene="ba245E14.1"
    /note="continues as ba245E14.1 in Em:AL358785
    match: proteins: Tr:Q92972 Tr:Q15774 Tr:Q07644 Sw:Q07231
    Sw:PI7028 Tr:O15535 Sw:Q15697 Tr:O14771 Tr:Q15636"
    /codon_start=1
    /evidence=not_experimental
    /product="d3874C20.1 (novel protein similar to zinc finger protein 306 (ZNF306, ZFP47))"
    /protein_id="CAI17573.1"
    /db_xref="GI:3169113"
    /db_xref="SPTREMBL:O60817"
    /translation="MARLESTALDAOSTEDOMELIVKVEEAGFPSSDLTSEG
    SHERFGFRYPAGRBAISRLREICROWLOPEHNSKQILELIVLOFTLEGNL
    GSWVRQHPESGEVYVLELYRLDEPAPVSGVCGOELCKMALTPAPSSOS
    SQPOLKALKHESVSGPQIDRVLYQVPLAHGCCREKRYASXLTPESQLKVED
    VALTPEWTFQDSSQGNLCRDEKQENHGSLSL"
misc_feature
    /complement(join(689..776,2604..2755,4087..4366))
    /gene="ba245E14.1"
    /note="match: GSS: Em:HSU77500"
    /complement(join(691..774,2606..2753,4089..4326))
    /gene="ba245E14.1"
    /note="match: GSS: Em:U77500"
    741..977
misc_feature
    /note="match: GSS: Em:U94424"
    741..880
misc_feature
    /note="match: GSS: Em:HSU94424"
    881..977
repeat_region
    /note="Aluub/FRAM repeat: matches 200..293 of consensus"
    1405..1448
repeat_region
    /note="4 copies 11 mer 88% conserved"
    1781..1950
repeat_region
    /note="L2 repeat: matches 2556..2741 of consensus"
    1983..2209
repeat_region
    /note="Aluub repeat: matches 7..230 of consensus"
    2216..2345
repeat_region
    /note="Aluub repeat: matches 191..311 of consensus"
    2810..3024
repeat_region
    /note="MIR repeat: matches 3..233 of consensus"
    3523..4001
repeat_region
    /note="Charlieta repeat: matches 1..508 of consensus"
    complement(4372..4551)
    /gene="ba245E14.1"
    /note="match: GSS: Em:HSU94428 Em:U94428"
    4976..5154
repeat_region
    /note="MLTIE repeat: matches 325..507 of consensus"
    5155..5471
repeat_region
    /note="Alusx repeat: matches 1..312 of consensus"
    5472..5534
repeat_region
    /note="MIRIE repeat: matches 507..568 of consensus"
    5533..5726
repeat_region
    /note="MER2 repeat: matches 122..345 of consensus"
    5728..5885
repeat_region
    /note="FRAM repeat: matches 0..154 of consensus"
    5887..6015
repeat_region
    /note="MER2 repeat: matches 1..130 of consensus"
    6321..6640
repeat_region
    /note="L2 repeat: matches 1156..1523 of consensus"
    6644..6827
repeat_region
    /note="MIR repeat: matches 10..234 of consensus"
    6828..7152
repeat_region
    /note="LTR18A repeat: matches 1..351 of consensus"
    LTR18A repeat: matches 1..351 of consensus"
    7153..7457
repeat_region
    /note="HERV18 repeat: matches 1..304 of consensus"
    HERV18 repeat: matches 1..304 of consensus"
    7440..7717
repeat_region
    /note="HERV18 repeat: matches 492..770 of consensus"
    HERV18 repeat: matches 492..770 of consensus"
    7716..7807
repeat_region
    /note="HERV18 repeat: matches 871..962 of consensus"
    HERV18 repeat: matches 871..962 of consensus"
    7795..8156
repeat_region
    /note="HERV18 repeat: matches 1390..1810 of consensus"
    HERV18 repeat: matches 1390..1810 of consensus"
    8156..8453
repeat_region
    /note="HERV18 repeat: matches 2119..2477 of consensus"
    HERV18 repeat: matches 2119..2477 of consensus"
    8428..9285
repeat_region
    /note="HERV18 repeat: matches 2718..3581 of consensus"
    HERV18 repeat: matches 2718..3581 of consensus"
    9286..9377
repeat_region
    /note="HERV18 repeat: matches 4394..4486 of consensus"
    HERV18 repeat: matches 4394..4486 of consensus"
    9682..10212
repeat_region
    /note="LTR18B repeat: matches 1..603 of consensus"
    LTR18B repeat: matches 1..603 of consensus"
    10268..10420
repeat_region
    /note="L2 repeat: matches 1553..1700 of consensus"
    10811..11123
repeat_region
    /note="AlusC repeat: matches 1..307 of consensus"
    11446..11530
repeat_region
    /note="MER5B repeat: matches 5..62 of consensus"
    11856..12304
repeat_region
    /note="LIMC5 repeat: matches 62..171 of consensus"
    11753..11852
repeat_region
    /note="Alusx repeat: matches 3..228 of consensus"
    12311..12466
repeat_region
    /note="LIMD3 repeat: matches 7375..7793 of consensus"
    12315..12480
repeat_region
    /note="26 copies 6 mer gagaga 59% conserved"
    12469..12611
repeat_region
    /note="83 copies 2 mer ga 59% conserved"
    12909..13039
repeat_region
    /note="LIMC5 repeat: matches 7810..7929 of consensus"
    13461..13656
repeat_region
    /note="LIMD3 repeat: matches 6449..6591 of consensus"
    13961..14008
repeat_region
    /note="MIR repeat: matches 12..210 of consensus"
    14753..14926
repeat_region
    /note="24 copies 2 mer ca 83% conserved"
    14787..15005
repeat_region
    /note="match: GSS: Em:AQ513873"
    14787..14983
repeat_region
    /note="match: GSS: Em:AQ542521"
    14811..15005
repeat_region
    /note="match: GSS: Em:AQ542442"
    14812..14994
repeat_region
    /note="match: GSS: Em:AQ366952"
    14812..14994
repeat_region
    /note="match: GSS: Em:AQ541753"
    complement(14813..15045)
repeat_region
    /note="match: GSS: Em:873060"
    complement(14813..14991)
repeat_region
    /note="match: GSS: Em:AQ082589"
```



```

misc_feature      complement(14814..14936)
                  /note="match: GSS: Em:AQ070409"
misc_feature      14817..15021
                  /note="match: GSS: Em:AQ543269"
misc_feature      /note="match: GSS: Em:AQ543269"
misc_feature      complement(14823..15015)
                  /note="match: GSS: Em:AQ729719"
misc_feature      complement(14833..15045)
                  /note="match: GSS: Em:AQ230792"
misc_feature      14833..14937
                  /note="match: GSS: Em:HS203XG5"

Query Match      1.18; Score 23; DB 9; Length 97847;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGAAAAA 2032
Db 5185 CAATTAAGAAAAA 5163

RESULT 45
AC017637 109151 bp DNA linear HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION AC017637
VERSION AC017637.1 GI:6554360
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 109151)
AUTHORS Adams,M. and Venter,J.C.
JOURNAL Direct Submission
Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211565 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Source 1..109151
Location/Qualifiers
  .organism="Drosophila melanogaster"
  .db_xref="taxon:7227"

BASE COUNT 31168 a 24117 c 24199 g 29667 t

Origin

Query Match      1.18; Score 23; DB 2; Length 109151;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGAAAAA 2032
Db 88267 CAATTAAGAAAAA 88289

RESULT 46
AC128720 111669 bp DNA linear HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-111D9, *** SEQUENCING IN PROGRESS ***
ACCESSION AC128720
VERSION AC128720.1 GI:21930183
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 111669)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrooks,S.L., Amarantunga,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowe,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burke,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathore,S.R., David,R.,
Dayle,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hoogson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,I.F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenberg,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,K.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,S., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Pearty,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoxan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scheper,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,P., Telford,B., Thomas,N., Thomas,S.,
Tsamis,K., Vasquez,L., Vera,V., Villison,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 111669)
Worley,K.C.
JOURNAL
REFERENCE
TITLE
JOURNAL
AUTHORS
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GSOI
Center clone name: CH230-111D9
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 45181 bases at least Q40
Consensus quality: 49519 bases at least Q30
Consensus quality: 52605 bases at least Q20
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1022: contig of 1022 bp in length  
 \* 1023 1122: gap of unknown length  
 \* 1123 2588: contig of 1466 bp in length  
 \* 2589 4308: gap of unknown length  
 \* 4309 4408: contig of 1620 bp in length  
 \* 4409 5928: gap of unknown length  
 \* 5929 6028: contig of 1520 bp in length  
 \* 6029 7515: gap of unknown length  
 \* 7516 8620: contig of 1487 bp in length  
 \* 8621 8720: gap of unknown length  
 \* 8721 9778: contig of 1058 bp in length  
 \* 9779 9878: gap of unknown length  
 \* 9879 11361: contig of 1483 bp in length  
 \* 11362 11461: gap of unknown length  
 \* 11462 13046: contig of 1585 bp in length  
 \* 13047 13146: gap of unknown length  
 \* 13147 14209: contig of 1063 bp in length  
 \* 14210 14309: gap of unknown length  
 \* 14310 15521: contig of 1212 bp in length  
 \* 15522 15621: gap of unknown length  
 \* 15622 17357: contig of 1736 bp in length  
 \* 17358 17457: gap of unknown length  
 \* 17458 18692: contig of 1235 bp in length  
 \* 18693 18792: gap of unknown length  
 \* 18793 19964: contig of 1172 bp in length  
 \* 19965 20064: gap of unknown length  
 \* 20065 21077: contig of 1013 bp in length  
 \* 21078 21177: gap of unknown length  
 \* 21178 22805: contig of 1628 bp in length  
 \* 22806 22905: gap of unknown length  
 \* 22906 24065: contig of 1160 bp in length  
 \* 24066 24165: gap of unknown length  
 \* 24166 25289: contig of 1134 bp in length  
 \* 25290 25399: gap of unknown length  
 \* 25400 26626: contig of 1227 bp in length  
 \* 26627 26726: gap of unknown length  
 \* 26727 27816: contig of 1090 bp in length  
 \* 27817 27916: gap of unknown length  
 \* 27917 29430: contig of 1514 bp in length  
 \* 29431 29530: gap of unknown length  
 \* 29531 31224: contig of 1694 bp in length  
 \* 31225 31324: gap of unknown length  
 \* 31325 33253: contig of 1929 bp in length  
 \* 33254 33353: gap of unknown length  
 \* 33354 35155: contig of 1802 bp in length  
 \* 35156 35255: gap of unknown length  
 \* 35256 36441: contig of 1186 bp in length  
 \* 36442 36541: gap of unknown length  
 \* 36542 37601: contig of 1060 bp in length  
 \* 37602 39355: contig of 1654 bp in length  
 \* 39356 39455: gap of unknown length  
 \* 39456 40916: contig of 1461 bp in length  
 \* 40917 41016: gap of unknown length  
 \* 41017 42467: contig of 1451 bp in length  
 \* 42468 42567: gap of unknown length  
 \* 42569 44379: contig of 1812 bp in length  
 \* 44380 44479: gap of unknown length  
 \* 44480 45520: contig of 1041 bp in length  
 \* 45521 45620: gap of unknown length  
 \* 45621 47184: contig of 1574 bp in length  
 \* 47185 47284: gap of unknown length  
 \* 47285 49498: contig of 2204 bp in length  
 \* 49499 49598: gap of unknown length  
 \* 49599 51450: contig of 1852 bp in length  
 \* 51451 51550: gap of unknown length  
 \* 51551 53556: contig of 2006 bp in length  
 \* 53557 53656: gap of unknown length  
 \* 53657 55732: contig of 2076 bp in length

\* 55832: gap of unknown length  
 \* 55833 58094: contig of 2262 bp in length  
 \* 58095 58194: gap of unknown length  
 \* 58195 59938: contig of 1744 bp in length  
 \* 59939 60038: gap of unknown length  
 \* 60039 61474: contig of 1436 bp in length  
 \* 61475 61574: gap of unknown length  
 \* 61575 63120: contig of 1546 bp in length  
 \* 63121 63220: gap of unknown length  
 \* 63221 64264: contig of 1044 bp in length  
 \* 64265 64364: gap of unknown length  
 \* 64365 66379: contig of 2015 bp in length  
 \* 66380 66479: gap of unknown length  
 \* 66480 67753: contig of 1274 bp in length  
 \* 67754 67853: gap of unknown length  
 \* 67854 69536: contig of 1683 bp in length  
 \* 69537 69636: gap of unknown length  
 \* 69637 71678: contig of 2042 bp in length  
 \* 71679 71778: gap of unknown length  
 \* 71779 73098: contig of 1320 bp in length  
 \* 73099 73198: gap of unknown length  
 \* 73199 75335: contig of 2137 bp in length  
 \* 75336 75435: gap of unknown length  
 \* 75436 76737: contig of 1302 bp in length  
 \* 76738 76837: gap of unknown length  
 \* 76839 78608: contig of 1771 bp in length  
 \* 78609 81516: contig of 2808 bp in length  
 \* 81517 81616: gap of unknown length  
 \* 81617 83645: contig of 2029 bp in length  
 \* 83646 83745: gap of unknown length  
 \* 83746 85904: contig of 2159 bp in length  
 \* 85905 86004: gap of unknown length  
 \* 86005 88523: contig of 2519 bp in length  
 \* 88524 88623: gap of unknown length  
 \* 88624 90880: contig of 2257 bp in length  
 \* 90881 90980: gap of unknown length  
 \* 90981 93536: contig of 2556 bp in length  
 \* 93537 93636: gap of unknown length  
 \* 93637 95861: contig of 2225 bp in length  
 \* 95862 95961: gap of unknown length  
 \* 95962 97882: contig of 1921 bp in length  
 \* 97883 97982: gap of unknown length

Query Match 1.1%; Score 23; DB 2; Length 111669;  
 Best Local Similarity 100.0%; Pred. No. 0.47;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 591 CCTGTGACATCATCCACAGAT 613  
 Db 65317 CCTGTGACATCATCCACAGAT 65339

RESULT 47  
 AC095664 133902 bp DNA linear HTG 11-JUL-2002  
 Rattus norvegicus clone CH230-8L17, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 LOCUS 68 uncloned pieces.

ACCESSION AC095664  
 VERSION AC095664.3 GI:21722631  
 KEYWORDS HTG; PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 133902)  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
 Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Bireva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Ellis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frattini, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoque, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratoch, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loughheed, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohapatra, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G., Ogunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulles, M., Ren, Y., Rives, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonakke, I., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telifod, B., Thomas, N., Thomas, S., Uman, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished  
2 (bases 1 to 133902)

Worley, K.C.  
Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 133902)

Worley, K.C.  
Direct Submission  
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence version replaced gi:17942230.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc.help@bcm.tmc.edu](mailto:hgsc.help@bcm.tmc.edu)  
Project Information  
Center project name: GCYM  
Center clone name: CH230-8L17  
----- Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap, version 0.990329  
Consensus quality: 7885 bases at least Q40  
Consensus quality: 86795 bases at least Q30  
Consensus quality: 92571 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 66 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

as soon as it is available and the accession number will  
\* be preserved.

1	1077	contig of 1077 bp in length
1078	1177	gap of unknown length
1178	2277	contig of 1100 bp in length
2278	2378	gap of unknown length
2378	3709	contig of 1331 bp in length
3709	3808	gap of unknown length
3809	4889	contig of 1081 bp in length
4890	4989	gap of unknown length
4990	6664	contig of 1674 bp in length
6664	6764	gap of unknown length
6764	8020	contig of 126 bp in length
8020	8119	gap of unknown length
8120	9702	contig of 1553 bp in length
9703	9802	gap of unknown length
9803	11676	contig of 1874 bp in length
11677	11776	gap of unknown length
11777	13246	contig of 1470 bp in length
13247	13346	gap of unknown length
13347	14543	contig of 1197 bp in length
14544	14643	gap of unknown length
14644	15890	contig of 1247 bp in length
15891	15990	gap of unknown length
15991	17240	contig of 1280 bp in length
17241	17340	gap of unknown length
17341	19252	contig of 1912 bp in length
19253	19352	gap of unknown length
19353	20819	contig of 1467 bp in length
20820	20919	gap of unknown length
20920	23340	contig of 2421 bp in length
23341	23440	gap of unknown length
23441	24900	contig of 1460 bp in length
24901	25000	gap of unknown length
25001	26021	contig of 1021 bp in length
26022	26121	gap of unknown length
26122	27281	contig of 1160 bp in length
27282	27381	gap of unknown length
27382	28383	contig of 1002 bp in length
28384	28483	gap of unknown length
28484	30162	contig of 1679 bp in length
30163	30262	gap of unknown length
30263	32073	contig of 1811 bp in length
32074	32173	gap of unknown length
32174	33657	contig of 1484 bp in length
33658	33757	gap of unknown length
33758	36013	contig of 2256 bp in length
36014	36113	gap of unknown length
36114	37587	contig of 1474 bp in length
37588	37687	gap of unknown length
37688	38690	contig of 1003 bp in length
38691	38790	gap of unknown length
38791	40372	contig of 1582 bp in length
40373	40472	gap of unknown length
40473	41611	contig of 1139 bp in length
41612	41711	gap of unknown length
41712	42841	contig of 1130 bp in length
42842	42941	gap of unknown length
42942	44303	contig of 1362 bp in length
44304	44403	gap of unknown length
44404	46227	contig of 1824 bp in length
46228	46327	gap of unknown length
46328	47528	contig of 1201 bp in length
47529	47628	gap of unknown length
47629	49306	contig of 1678 bp in length
49307	49406	gap of unknown length
49407	50779	contig of 1373 bp in length
50780	50879	gap of unknown length
50880	52168	contig of 1269 bp in length
52169	52268	gap of unknown length
52269	53781	contig of 1513 bp in length
53782	53882	gap of unknown length
53882	55307	contig of 1426 bp in length

\* 55308 55407: gap of unknown length  
 \* 55408 56741: contig of 1334 bp in length  
 \* 56742 56841: gap of unknown length  
 \* 56842 58327: contig of 1486 bp in length  
 \* 58328 58427: gap of unknown length  
 \* 58428 60589: contig of 2162 bp in length  
 \* 60590 60689: gap of unknown length  
 \* 60690 62255: contig of 1566 bp in length  
 \* 62256 62355: gap of unknown length  
 \* 62356 64065: contig of 1710 bp in length  
 \* 64066 64165: gap of unknown length  
 \* 64166 65421: contig of 1256 bp in length  
 \* 65422 67166: gap of unknown length  
 \* 67167 67266: gap of unknown length  
 \* 67267 68899: contig of 1633 bp in length  
 \* 68900 71156: gap of unknown length  
 \* 71157 71256: gap of unknown length  
 \* 71257 73078: contig of 1822 bp in length  
 \* 73079 73178: gap of unknown length  
 \* 73179 74887: contig of 1709 bp in length  
 \* 74888 74987: gap of unknown length  
 \* 74988 77893: contig of 2906 bp in length  
 \* 77894 77993: gap of unknown length  
 \* 77994 79712: contig of 1718 bp in length  
 \* 79712 82367: gap of unknown length  
 \* 82368 82467: gap of unknown length  
 \* 82468 84455: contig of 1988 bp in length  
 \* 84456 84555: gap of unknown length  
 \* 84556 85795: contig of 1240 bp in length  
 \* 85796 85895: gap of unknown length  
 \* 85896 87801: contig of 1906 bp in length  
 \* 87802 87901: gap of unknown length  
 \* 87902 89579: contig of 1678 bp in length

Query Match 1.1%; Score 23; DB 2; Length 133902;  
 Best Local Similarity 100.0%; Pred. No. 0.47;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 GTGCTCAAGAGGTGCGCTCTT 726  
 Db 87153 GTGCTCAAGAGGTGCGCTCTT 87175

RESULT 48  
 AC116962 141017 bp DNA linear HTG 04-APR-2002  
 LOCUS Dictyostelium discoideum chromosome 2 map 1775121-1916136 strain  
 DEFINITION A44, \*\*\* SEQUENCING IN PROGRESS \*\*\*; in ordered pieces.  
 ACCESSION AC116962.1 GI:19920061  
 VERSION HTG: HTGS\_PHASE2.  
 KEYWORDS Dictyostelium discoideum.  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 REFERENCE 1 (bases 1 to 141017)  
 Lehnman, R., Baumgart, C., Parra, G., April, J. F., Guiso, R., Kumpf, R.,  
 Jungel, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and  
 Noegel, A. A.  
 Sequence and Analysis of Chromosome 2 of Dictyostelium  
 Unpublished  
 JOURNAL The Dictyostelium Genome Sequencing Consortium  
 REMARK 2 (bases 1 to 141017)  
 AUTHORS Baumgart, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-APR-2002) Genome Analysis, Institute of Molecular  
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
 COMMENT CDS predictions from GenBank may contain errors. Further information  
 is available from IMB Jena, Department of Genome Analysis  
 (http://genome.imb-jena.de/dictyostelium/)

and the University Cologne, Institute for Biochemistry I  
 (http://www.uni-koeln.de/dictyostelium/project.shtml  
 Funding  
 Agency : Deutsche Forschungsgemeinschaft (DFG).  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source  
 1. 141017  
 /organism="Dictyostelium discoideum"  
 /strain="A44"  
 /db\_xref="taxon:44689"  
 /chromosome="2"  
 /map="1775121-1916136"  
 BASE COUNT 55240 a 15334 c 15477 g 54866 t 100 others  
 ORIGIN

Query Match 1.1%; Score 23; DB 2; Length 141017;  
 Best Local Similarity 100.0%; Pred. No. 0.48;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032  
 Db 28316 CAATATAGAAAAA 28338

RESULT 49  
 AC099262/c 143351 bp DNA linear HTG 13-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-107H13, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 82 unordered pieces.  
 ACCESSION AC099262.4 GI:21731147  
 VERSION HTG: HTGS\_PHASE1.  
 KEYWORDS Rattus norvegicus  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 1 (bases 1 to 143351)  
 Muzny, D. M., Adams, C., Adio-Oduola, B., Alt-Osman, F. R., Allen, C.,  
 Alsbrooks, S. L., Amaral-Tunget, H. C., Are, J. R., Ayala, M., Banks, T.,  
 Barbarella, J., Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P.,  
 Bunay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C.,  
 Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C. D., Cox, C., Coyle, M. D., Dethorne, S. R., David, R.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A.,  
 Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H.,  
 Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,  
 Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escoto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Housli, F., Howard, S., Huber, J., Huijy, S., Hume, J., Jackson, L. E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvay, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lonsleged, H.,  
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maneshari, M., Mapue, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenko, S., Ogun, M., Okunnu, G.,  
 Orangunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojubo-Kan, I., Rolfe, M., Ruliz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,

Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, S.,  
Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, C.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, D., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

\*\*\*\*\*  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
\*\*\*\*\*

Unpublished  
2 (bases 1 to 143351)  
Worley, K.C.  
Direct Submission  
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 143351)  
Worley, K.C.  
Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gi:17974677.

\*\*\*\*\*  
Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GJXX  
Center clone name: CH230-107H13  
----- Summary Statistics  
Sequencing vector: plasmid,  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 77977 bases at least Q40  
Consensus quality: 86348 bases at least Q20  
Consensus quality: 92370 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 82 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1041: contig of 1041 bp in length  
1042 1141: gap of unknown length  
1043 2212: contig of 1071 bp in length  
1044 2213: gap of unknown length  
1045 2313: contig of 1029 bp in length  
1046 3341: gap of unknown length  
1047 3441: gap of unknown length  
1048 4442: contig of 1001 bp in length  
1049 4443: gap of unknown length  
1050 4543: gap of unknown length  
1051 5662: contig of 1120 bp in length  
1052 5663: gap of unknown length  
1053 5762: gap of unknown length  
1054 7844: contig of 2082 bp in length  
1055 7845: gap of unknown length  
1056 7945: gap of unknown length  
1057 9672: contig of 1728 bp in length  
1058 9673: gap of unknown length  
1059 9772: gap of unknown length  
1060 10810: contig of 1038 bp in length  
1061 10811: gap of unknown length  
1062 10910: contig of 1012 bp in length  
1063 11922: contig of 1012 bp in length  
1064 11923: gap of unknown length  
1065 12022: contig of 1015 bp in length  
1066 13037: gap of unknown length  
1067 13038: gap of unknown length  
1068 13137: gap of unknown length  
1069 13138: gap of unknown length  
1070 14230: contig of 1093 bp in length  
1071 14231: gap of unknown length  
1072 14331: contig of 1145 bp in length  
1073 14331: gap of unknown length  
1074 15475: contig of 1145 bp in length  
1075 15476: gap of unknown length  
1076 15575: gap of unknown length  
1077 15576: gap of unknown length  
1078 16609: contig of 1034 bp in length  
1079 16610: gap of unknown length  
1080 16709: contig of 1728 bp in length  
1081 18437: gap of unknown length  
1082 18438: gap of unknown length  
1083 18537: gap of unknown length  
1084 18538: gap of unknown length  
1085 19554: contig of 1027 bp in length  
1086 19564: gap of unknown length  
1087 19665: gap of unknown length  
1088 19665: gap of unknown length  
1089 20756: contig of 1092 bp in length  
1090 20757: gap of unknown length  
1091 20857: gap of unknown length  
1092 22151: contig of 1294 bp in length  
1093 22151: gap of unknown length  
1094 22251: contig of 1082 bp in length  
1095 22333: gap of unknown length  
1096 22333: gap of unknown length  
1097 22477: contig of 1315 bp in length  
1098 22478: gap of unknown length  
1099 24447: gap of unknown length  
1100 24848: contig of 1023 bp in length  
1101 25870: gap of unknown length  
1102 25871: gap of unknown length  
1103 25971: contig of 1291 bp in length  
1104 27261: gap of unknown length  
1105 27262: gap of unknown length  
1106 27362: gap of unknown length  
1107 28491: gap of unknown length  
1108 28590: gap of unknown length  
1109 28591: gap of unknown length  
1110 30486: contig of 1896 bp in length  
1111 30487: gap of unknown length  
1112 30586: gap of unknown length  
1113 32080: contig of 1494 bp in length  
1114 32081: gap of unknown length  
1115 32181: gap of unknown length  
1116 33181: gap of unknown length  
1117 33185: contig of 1205 bp in length  
1118 33186: gap of unknown length  
1119 33485: gap of unknown length  
1120 33486: gap of unknown length  
1121 34600: contig of 1115 bp in length  
1122 34700: gap of unknown length  
1123 34701: contig of 1893 bp in length  
1124 36593: gap of unknown length  
1125 36594: gap of unknown length  
1126 37762: contig of 1066 bp in length  
1127 37763: gap of unknown length  
1128 37862: gap of unknown length  
1129 39209: gap of unknown length  
1130 39210: gap of unknown length  
1131 39309: gap of unknown length  
1132 40939: contig of 1630 bp in length  
1133 40940: gap of unknown length  
1134 41039: gap of unknown length  
1135 42390: contig of 1351 bp in length  
1136 42391: gap of unknown length  
1137 42490: gap of unknown length  
1138 42491: gap of unknown length  
1139 43515: contig of 1025 bp in length  
1140 43516: gap of unknown length  
1141 43616: gap of unknown length  
1142 44805: contig of 1190 bp in length  
1143 44806: gap of unknown length  
1144 44905: gap of unknown length  
1145 44906: contig of 1861 bp in length  
1146 46765: gap of unknown length  
1147 46767: gap of unknown length  
1148 46866: gap of unknown length  
1149 48277: contig of 1411 bp in length  
1150 48278: gap of unknown length  
1151 48377: gap of unknown length  
1152 48378: gap of unknown length  
1153 49729: contig of 1352 bp in length  
1154 49829: gap of unknown length  
1155 49830: gap of unknown length  
1156 51737: contig of 1908 bp in length  
1157 51738: gap of unknown length  
1158 51837: gap of unknown length  
1159 53416: contig of 1579 bp in length  
1160 53417: gap of unknown length  
1161 53516: gap of unknown length  
1162 53517: gap of unknown length  
1163 55685: contig of 2165 bp in length  
1164 55686: gap of unknown length  
1165 55785: gap of unknown length  
1166 55786: gap of unknown length  
1167 57414: contig of 1629 bp in length  
1168 57514: gap of unknown length  
1169 57515: gap of unknown length  
1170 58773: contig of 1255 bp in length  
1171 58774: gap of unknown length  
1172 58873: gap of unknown length  
1173 60454: contig of 1581 bp in length  
1174 60455: gap of unknown length  
1175 60555: gap of unknown length  
1176 60556: gap of unknown length  
1177 61862: contig of 1306 bp in length  
1178 61863: gap of unknown length  
1179 61865: gap of unknown length  
1180 63273: contig of 1311 bp in length  
1181 63274: gap of unknown length  
1182 63374: gap of unknown length  
1183 65180: contig of 1807 bp in length  
1184 65181: gap of unknown length  
1185 65280: gap of unknown length  
1186 65281: gap of unknown length  
1187 66544: contig of 1264 bp in length  
1188 66545: gap of unknown length  
1189 66645: gap of unknown length  
1190 66646: gap of unknown length  
1191 67954: gap of unknown length  
1192 68054: gap of unknown length  
1193 68055: contig of 1121 bp in length  
1194 69175: gap of unknown length  
1195 69176: gap of unknown length  
1196 69275: gap of unknown length  
1197 71084: contig of 1809 bp in length

\* 71085 71184: gap of unknown length  
 \* 72535 72535: contig of 1351 bp in length  
 \* 72536 72635: gap of unknown length  
 \* 72636 73833: contig of 1198 bp in length  
 \* 73833 73933: gap of unknown length  
 \* 73933 75767: contig of 1834 bp in length  
 \* 75768 75867: gap of unknown length  
 \* 75868 77808: contig of 1941 bp in length  
 \* 77809 77908: gap of unknown length  
 \* 77909 77921: contig of 1813 bp in length

Query Match 1.1%; Score 23; DB 2; Length 143351;  
 Best Local Similarity 100.0%; Pred. No. 0.48;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 572 GCCCTGCTCTGCACCTGCCTG 594  
 Db 132644 GCCCTGCTCTGCACCTGCCTG 132622

RESULT 50  
 AC115914 159170 bp DNA linear HTG 20-JUN-2002  
 LOCUS AC115914  
 DEFINITION Mus musculus clone RP24-487J5, WORKING DRAFT SEQUENCE, 16 ordered pieces.  
 AC115914 GI:21490457  
 VERSION HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 159170)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 JOURNAL Mus musculus, clone RP24-487J5  
 REFERENCE 2 (bases 1 to 159170)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhaltier, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillio, J., Vassiliev, H., Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 159170)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhaltier, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillio, J., Vassiliev, H., Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 159170)

COMMENT  
 TITLE  
 JOURNAL  
 Submitted (20-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 20, 2002 this sequence version replaced g1.19683593.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information -----  
 Center project name: 124872  
 Center clone name: 487.J.5  
 ----- Summary Statistics -----  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 154428 bases at least Q40  
 Consensus quality: 156177 bases at least Q30  
 Consensus quality: 156924 bases at least Q20  
 Insert size: 163000; agarose-ff  
 Insert size: 157670; sum-of-contigs  
 Quality coverage: 6.4 in Q20 bases; sum-of-contigs  
 Quality coverage: 6.6 in Q20 bases; sum-of-contigs  
 ----- NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved. -----  
 1 1032: contig of 1032 bp in length  
 \* 1033 1132: gap of 100 bp  
 \* 1133 4008: contig of 2876 bp in length  
 \* 4009 4108: gap of 100 bp  
 \* 4109 6871: contig of 2763 bp in length  
 \* 6872 6971: gap of 100 bp  
 \* 6972 8935: contig of 1964 bp in length  
 \* 8936 9035: gap of 100 bp  
 \* 9036 13009: contig of 3974 bp in length  
 \* 13010 13109: gap of 100 bp  
 \* 13110 15598: contig of 2489 bp in length  
 \* 15599 15698: gap of 100 bp  
 \* 15699 19053: contig of 3355 bp in length  
 \* 19054 19153: gap of 100 bp  
 \* 19154 26269: contig of 7116 bp in length  
 \* 26270 26369: gap of 100 bp  
 \* 26370 34645: contig of 8276 bp in length  
 \* 34646 34745: gap of 100 bp  
 \* 34746 46455: contig of 11710 bp in length  
 \* 46456 46555: gap of 100 bp  
 \* 46556 55127: contig of 8872 bp in length  
 \* 55128 55227: gap of 100 bp  
 \* 55228 67686: contig of 12455 bp in length  
 \* 67687 67786: gap of 100 bp  
 \* 67787 90042: contig of 22256 bp in length

*	90043	90142: gap of	100 bp
*	90143	109818: contig of	19676 bp in length
*	109819	109918: gap of	100 bp
*	105919	130304: contig of	20386 bp in length
*	130305	130404: gap of	100 bp
*	130405	159170: contig of	28766 bp in length.

## FEATURES

Source

```
misc_feature      1..1032 /note="assembly_fragment"  
misc_feature     1133..4008 /note="assembly_fragment"  
misc_feature     4109..6871 /note="assembly_fragment"  
misc_feature     6972..8935 /note="assembly_fragment"  
misc_feature     9036..13009 /note="assembly_fragment"  
misc_feature     13110..15598 /note="assembly_fragment"  
misc_feature     15699..18053 /note="assembly_fragment"  
misc_feature     19154..26269 /note="assembly_fragment"  
misc_feature     26370..34645 /note="assembly_fragment"  
misc_feature     34746..46455 /note="assembly_fragment"  
misc_feature     46556..55127 /note="assembly_fragment"  
misc_feature     55228..67686 /note="assembly_fragment"  
misc_feature     67787..90042 /note="assembly_fragment"  
misc_feature     90143..109818 /note="assembly_fragment"  
misc_feature     109919..130304 /note="assembly_fragment"  
misc_feature     130405..159170 /note="assembly_fragment"  
misc_feature     159171..18053 /note="assembly_fragment"  
  
BASE COUNT      41889 a 35541 c 36173 g 44061 t 1506 others  
BRIGIN
```

Query Match	1.1%	Score 23	DB 2	Length 159170
Best Local Similarity	100.0%	Pred. No. 0.48		
Matches 23	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	704	GTGCTCAAGGAGGTCGCTTCTT	726	
Db 145557	GTGCTCAAGGAGGTCGCTTCTT	145579		

Search completed: January 11, 2003, 02:09:57  
Job time : 11357 secs

**This Page Blank (uspto)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:56:19 ; Search time 427 Seconds  
(without alignments)  
10716.770 Million cell updates/sec

Title: US-09-816-825-1  
Perfect score: 2032  
Sequence: 1 ggcctcagagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO\_NUC  
Gapop 60.0 ; Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 20

Total number of hits satisfying chosen parameters: 145

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :  
1: N\_Geneseq\_101002:\*  
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2032	100.0	2032	20	AAZ20792 Human glycosyl sul
2	1802	88.7	1979	22	AAK94229 Human full-length
3	1615	79.5	2065	21	AAZ94211 Human transferrase
4	1211	59.6	1333	24	AA516947 Human L-selectin s
5	741	36.5	877	22	AAK91803 Human CDNA 5'-end
6	741	36.5	877	22	AAK93921 Human CDNA clone r
7	505	24.9	517	24	AAK54724 Human colon cancer
8	351	17.3	548	22	AAK92588 Human CDNA 3'-end
9	306	15.1	2988	21	AAK61556 Human ORFX ORF1711

10	286	14.1	389	24	AAK54794 Human colon cancer
11	60	3.0	60	24	ABN38074 Human spliced tran
12	59	2.9	1647	24	AAD24670 Human drug metabol
13	59	2.9	1694	22	AAD02700 Human glycosyl sul
14	59	2.9	2044	22	AAD02699 Human glycosyl sul
15	59	2.9	2170	22	AAD02898 Human glycosyl sul
16	59	2.9	2544	24	ABN89506 Human corneal N-ac
17	59	2.9	48436	24	ABN89506 Human corneal N-ac
18	59	2.9	160552	22	AAD02697 Human glycosyl sul
19	37	1.8	30	20	AAZ20798 PCR primer for gly
20	30	1.5	30	24	ABK70821 Probe for human g
21	27	1.3	1926	20	AAZ20793 Mouse glycosyl sul
22	23	1.1	616	21	AAK74310 Human secreted pro
23	23	1.1	1142	22	AAK41008 cDNA encoding nove
24	23	1.1	1937	24	AAK16948 Murine intestinal-
25	23	1.1	1989	22	AAD02696 Mouse glycosyl sul
26	23	1.1	2411	23	ABV29163 Human prostate exp
27	23	1.1	2860	18	AAK63303 A. chrysogenum O-a
28	23	1.1	4871	23	ABK16362 Drosophila melanog
29	23	1.1	5101	23	ABK16364 Drosophila melanog
30	23	1.1	10138	23	ABK16360 Drosophila melanog
31	22	1.1	22	24	AAK70834 PCR primer for hum
32	22	1.1	22	24	AAK16960 Human L-selectin s
33	22	1.1	22	24	AAK16962 Human L-selectin s
34	22	1.1	22	24	AAK16964 Human L-selectin s
35	22	1.1	341	23	ABV38642 Human prostate exp
36	22	1.1	502	23	ABV34232 Human prostate exp
37	22	1.1	1781	22	AAK41556 cDNA encoding nove
38	22	1.1	1782	24	ABK63555 Rat sequence diffe
39	22	1.1	2415	22	AAH75633 Human ribosomal pr
40	22	1.1	3138	12	AAO11712 Shuttle vector PMU
41	22	1.1	3597	22	ABA82684 HSM800936 gene SFO
42	22	1.1	73947	23	ABK07230 Drosophila melanog
43	21	1.0	21	24	AAK10835 PCR primer for hum
44	21	1.0	21	24	AAK10835 Human L-selectin s
45	21	1.0	51	22	AAK16961 Human SNP oligonuc
46	21	1.0	262	23	AAV08745 Human prostate exp
47	21	1.0	294	22	AAK63356 DNA encoding larva
48	21	1.0	397	23	AAK56821 Human prostate exp
49	21	1.0	405	22	AAK92205 Human prostate exp
50	21	1.0	411	24	ABN94490 Human polynucleoti
51	21	1.0	411	24	ABN94490 Gene #1438 used to
52	21	1.0	428	23	ABK62995 Breast cancer rela
53	21	1.0	458	23	AAK83335 Human prostate exp
54	21	1.0	733	22	AAK196611 DNA encoding novel
55	21	1.0	785	24	AAK960921 Human neuroblastom
56	21	1.0	986	24	ABN98495 Arabidopsis thalia
57	21	1.0	998	21	AAK77903 Human cancer assoc
58	21	1.0	1563	24	ABO73783 Human colon specifi
59	21	1.0	1461	21	AAK78402 Human secreted pro
60	21	1.0	1469	24	AAK10276 Human cDNA encodin
61	21	1.0	1493	11	AAO04492 Sequence encoding
62	21	1.0	1493	20	AAK33994 Human osteopontin
63	21	1.0	1792	21	AAK77865 Human cancer assoc
64	21	1.0	1792	22	AAK72777 Human prostate can
65	21	1.0	2013	21	AAK78178 Human cancer assoc
66	21	1.0	2655	22	AAK72751 Human prostate can
67	21	1.0	2835	22	AAK52037 Human polynucleoti
68	21	1.0	6608	24	ABK34034 Human immune syst
69	21	1.0	10004	22	ABK14483 Human nervous syst
70	21	1.0	10246	22	ABK14485 Human nervous syst
71	20	1.0	20	22	AAK50713 Polyprymidine Crti
72	20	1.0	88	21	AAK25980 Human secreted pro
73	20	1.0	161	24	ABK186331 Human ovarian can
74	20	1.0	225	23	ABV57865 Human prostate exp
75	20	1.0	293	22	AAK12536 Human breast cance
76	20	1.0	295	16	AAO93982 Rat allograft infl
77	20	1.0	295	21	AAK65602 Rat allograft infl
78	20	1.0	308	22	AAK16693 Human breast cance
79	20	1.0	362	16	AAO99378 Rat allograft infl
80	20	1.0	362	21	AAK65598 Rat allograft infl
81	20	1.0	364	22	AAK65598 Human breast cance
82	20	1.0	384	23	ABK13665 Human prostate exp

83	20	1.0	387	20	AA40849	Secreted protein E
84	20	1.0	403	22	AA18736	Human polynucleoti
85	20	1.0	405	23	ABV34779	Human prostate exp
86	20	1.0	405	23	ABV34779	Human prostate exp
87	20	1.0	410	23	ABV18937	Human prostate exp
88	20	1.0	410	23	ABV18937	Human prostate exp
89	20	1.0	435	24	ABV56080	Human prostate exp
90	20	1.0	451	23	ABV48716	Human prostate exp
91	20	1.0	468	22	ABV41497	Human nervous syst
92	20	1.0	474	22	AA560615	Human cancer agent
93	20	1.0	488	22	AA529035	CDNA encoding for
94	20	1.0	517	23	ABR62399	Rat sequence diffie
95	20	1.0	527	24	ABV58675	Human prostate exp
96	20	1.0	554	24	ABK54892	Human colon cancer
97	20	1.0	553	24	AA113827	Human breast cancer
98	20	1.0	630	21	AA116113	Human prostate can
99	20	1.0	804	22	AA197306	Human neuroblastom
100	20	1.0	873	22	AA122693	Human breast cancer
101	20	1.0	977	21	AA264465	Human prostate can
102	20	1.0	986	21	ABQ34314	Oligonucleotide fo
103	20	1.0	986	24	ABQ34315	Oligonucleotide fo
104	20	1.0	1094	24	AA596308	Arabidopsis CDNA e
105	20	1.0	1104	24	AA580708	Soybean chlorophyl
106	20	1.0	1117	21	AA118221	Lung cancer associ
107	20	1.0	1243	21	AAQ3137	Arabidopsis thalia
108	20	1.0	1297	13	AAQ21645	3' coding sequence
109	20	1.0	1316	22	AAH72877	Human secreted pro
110	20	1.0	1389	21	AA339055	Human secreted pro
111	20	1.0	1504	21	AA38952	Arabidopsis thalia
112	20	1.0	1526	21	AACT6491	Human OREX ORE2046
113	20	1.0	1587	21	AA34626	Drosophila dm11 t
114	20	1.0	1708	21	AACT4428	Human secreted pro
115	20	1.0	1770	14	AAQ37684	P. falciparum antiq
116	20	1.0	1770	14	AAQ37684	P. falciparum antiq
117	20	1.0	1869	22	AA444878	Human INTERCEPT
118	20	1.0	1869	22	AA45014	Human secreted pro
119	20	1.0	1869	22	AA45015	Human secreted pro
120	20	1.0	1869	22	AA45016	Human secreted pro
121	20	1.0	1869	22	AA45017	Human secreted pro
122	20	1.0	2290	22	AA59490	Frog CDNA encoding
123	20	1.0	2429	22	AA73413	Grand fir monoterp
124	20	1.0	2441	24	AA562635	CDNA sequence #422
125	20	1.0	2843	24	ABR48708	CDNA encoding huma
126	20	1.0	2888	21	AA96483	CDNA encoding a hu
127	20	1.0	2894	20	AA86127	EST sequence for D
128	20	1.0	2959	24	AA596336	Arabidopsis CDNA e
129	20	1.0	3492	20	AA35745	CDNA encoding a pr
130	20	1.0	3515	22	AA75343	Human TGF-beta rec
131	20	1.0	3763	19	AA58194	Human myosin I-cha
132	20	1.0	3763	23	AA58194	DNA encoding novel
133	20	1.0	5814	24	ABN87855	Human ovary specif
134	20	1.0	7138	24	ABR28455	DNA transcription
135	20	1.0	8652	23	ABLO3026	Drosophila melanog
136	20	1.0	9516	22	AAK73555	Human immune/haema
137	20	1.0	9516	22	AAK81394	Human immune/haema
138	20	1.0	13192	22	AAK73556	Human immune/haema
139	20	1.0	13192	22	AAK81396	Human immune/haema
140	20	1.0	14482	23	ABLI0086	Drosophila melanog
141	20	1.0	20990	22	ABLO3102	Drosophila melanog
142	20	1.0	35959	22	AAK78275	Human immune/haema
143	20	1.0	38653	22	AA544513	Human LEX1 DNA cl
144	20	1.0	147419	22	ABR83574	Human CDNA differe
145	20	1.0	349980	22	ABR86431	Pyrococcus abyssi

## ALIGNMENTS

RESULT 1  
AA220792  
ID AA220792 standard: DNA: 2032 BP.  
XX  
AC AA220792:

XX	08-DEC-1999 (first entry)	
DT	Human glycosyl sulfotransferase-3 coding sequence.	
XX		
DE	Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;	
XX	selectin binding interaction; inflammation; lymphocyte homing; human;	
KW	secondary lymph organ; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	W09949018-A1.	
XX	30-SEP-1999.	
PD		
XX	26-FEB-1999; 99WO-US04316.	
XX		
PF	20-MAR-1998; 98US-0045284.	
XX		
PR	12-NOV-1998; 98US-0190911.	
XX		
PA	(REGC ) UNIV CALIFORNIA.	
XX	(SYNP ) SYNTX USA INC.	
PI	Bistrup A, Rosen SD, Tangemann K, Hemmerich S;	
XX		
DR	WPI: 1999-580442/49.	
XX	P-PSDB: AAY39918.	
PT	Human and murine glycosyl sulfotransferase 3 and related	
XX	polynucleotides	
PS	Claim 4; Fig 1; 59pp: English.	
XX		
CC	This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of	
CC	the invention. The nucleic acid sequences, probes and primers derived	
CC	from these, proteins and antibodies are useful in detecting homologues.	
CC	The sequences, antibodies and methods are useful in the diagnosis and	
CC	treatment of diseases associated with selectin binding interactions,	
CC	including conditions associated with or resulting from the homing of	
CC	leucocytes to sites of inflammation and the normal homing of lymphocytes	
CC	to secondary lymph organs.	
XX		
SQ	Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other:	
Query Match	100.0%; Score 2032; DB 20; Length 2032;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2032; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GGCTGAGGCCAGAGTGCCTCCAGTCTGGGGGAAATGCTTCTTATTTGCTTCCAG 60	
DB	1 GGCTGAGGCCAGAGTGCCTCCAGTCTGGGGGAAATGCTTCTTATTTGCTTCCAG 60	
QY	61 CCACCTCAGCAGAGTCCGCCACCCCTTGAGTCTCAGAGTGTAAAGCTGTACTTCA 120	
DB	61 CCACCTCAGCAGAGTCTCCGCCACCCCTTGAGTCTCAGAGTGTAAAGCTGTACTTCA 120	
QY	121 CAGCTTCTCGGAGGAGAGTCTTCTCAAGCCGCTTTCAGAGGCTTCCACTTCAGCAC 180	
DB	121 CAGCTTCTCGGAGGAGAGTCTTCTCAAGCCGCTTTCAGAGGCTTCCACTTCAGCAC 180	
QY	181 AATGCTACGCTTAAATAAATGAAGCTGCTGTTTGTGTTCCAGAGGCGCATCTT 240	
DB	181 AATGCTACGCTTAAATAAATGAAGCTGCTGTTTGTGTTCCAGAGGCGCATCTT 240	
QY	241 GGCTTATTTCTTCACATGTACAGCCACACATCAGTCCCTGTCTATGAAGGACACCC 300	
DB	241 GGCTTATTTCTTCACATGTACAGCCACACATCAGTCCCTGTCTATGAAGGACACCC 300	
QY	301 CGAGCGCATGACAGCTGCTGTTCTCTTCCCTGGGCGCTCTGCTTTTGTGGGCA 360	
DB	301 CGAGCGCATGACAGCTGCTGTTCTCTTCCCTGGGCGCTCTGCTTTTGTGGGCA 360	
QY	361 GCTTTTGGGCGAGCACCCAGATGTTTCTACGTATGAGCCCGGCTGACAGTGTGAT 420	

```
Db 361 GCTTTTGGGAGAGACCAGATGTTTCTTACCTGATGAGCCGCCCTGGACACCTGTGTGAT 420
QY 421 GACCTTCAGAGAGACACCCGCTGATGCTCAGATGGCTGTGGGGATGATGATGAGGC 480
Db 421 GACCTTCAGAGAGACACCCGCTGATGCTCAGATGGCTGTGGGGATGATGATGAGGC 480
QY 481 CGTCTTCTTGTGCGACATGACGCTCTTGTGATGCTTACATGGAACCTGTGCCGAGACA 540
Db 481 CGTCTTCTTGTGCGACATGACGCTCTTGTGATGCTTACATGGAACCTGTGCCGAGACA 540
QY 541 GTCCAGCCTCTTTCAGTGGGAGAGACGCCGCGCTGTGTCTCTGCTCAGCTGTGTGAT 600
Db 541 GTCCAGCCTCTTTCAGTGGGAGAGAGACGCCGCGCTGTGTCTCTGCTCAGCTGTGACAT 600
QY 601 CATCCACACAGATGAAATCATCCGCCGCGCTCAGCTGAGGCTCCTGTGACATCAGACCC 660
Db 601 CATCCACACAGATGAAATCATCCGCCGCGCTCAGCTGAGGCTCCTGTGACATCAGACCC 660
QY 661 CTTTGAGTGGTGGAGAGAGGCTGCCCTCCTTACAGCCAGCTGGTGGTCAAGAGGTGGC 720
Db 661 CTTTGAGTGGTGGAGAGAGGCTGCCCTCCTTACAGCCAGCTGGTGGTCAAGAGGTGGC 720
QY 721 CTTTCTTCAACTGACAGTCCCTCTACCCGCTGCTGAAGACCCCTCCCTCAACCTGATAT 780
Db 721 CTTTCTTCAACTGACAGTCCCTCTACCCGCTGCTGAAGACCCCTCCCTCAACCTGATAT 780
QY 781 CGTGACACCTGTGCTCCGGGAGCCCGCGGCGCTGTCCGTTCCCGAGACGACACAAAGGAGA 840
Db 781 CGTGACACCTGTGCTCCGGGAGCCCGCGGCGCTGTCCGTTCCCGAGAGCGACAAAGGAGA 840
QY 841 TCTCATGATGATGACAGTGGCATTTGATGAGGAGACATGAGAGAGAAATCAAGAGAGAGA 900
Db 841 TCTCATGATGATGACAGTGGCATTTGATGAGGAGACATGAGAGAGAAATCAAGAGAGAGA 900
QY 901 CCAACCCCTACTATGTGATGACAGTGCATCTGCGAAGCCAGCTGAGATCTACAAAGACAT 960
Db 901 CCAACCCCTACTATGTGATGACAGTGCATCTGCGAAGCCAGCTGAGATCTACAAAGACAT 960
QY 961 CCAGTCCCTTGCCCAAGGCGCTGCGAGAGACGCTACTGCTGTGCGCTATGAGACCTGGC 1020
Db 961 CCAGTCCCTTGCCCAAGGCGCTGCGAGAGACGCTACTGCTGTGCGCTATGAGAGCCTGGC 1020
QY 1021 TCGAGCCCTGTGGCCAGACATCTCCGAAATGATGAATTCGTGGATGGAATCTTGGC 1080
Db 1021 TCGAGCCCTGTGGCCAGACATCTCCGAAATGATGAATTCGTGGATGGAATCTTGGC 1080
QY 1081 CCATCTTCAGACCTGGGTGCTAATCATACCCGAGGCAAGGGCATGGGTGACACGCTTT 1140
Db 1081 CCATCTTCAGACCTGGGTGCTAATCATACCCGAGGCAAGGGCATGGGTGACACGCTTT 1140
QY 1141 CCACACAATGCCAGGGATGCCCTTAATGTTCTCCAGGCTGGCGCTGCTTGGCCCTA 1200
Db 1141 CCACACAATGCCAGGGATGCCCTTAATGTTCTCCAGGCTGGCGCTGCTTGGCCCTA 1200
QY 1201 TGAAGAGGTTTCTGACTTCGAAAGCCTGTGGCATGCAATTTGCTGGGCTACCG 1260
Db 1201 TGAAGAGGTTTCTGACTTCGAAAGCCTGTGGCATGCAATTTGCTGGGCTACCG 1260
QY 1261 CCAGCTTCGATCTGACACAGAGAGAAACCTGTGCTGATTTCTGTCTACCTGAGC 1320
Db 1261 CCAGCTTCGATCTGACACAGAGAGAAACCTGTGCTGATTTCTGTCTACCTGAGC 1320
QY 1321 TGTCCCTGAGCAAAATCCACTAGAGGCTTGAAGAGCTTGTGCTGCCACTGTGTGACC 1380
Db 1321 TGTCCCTGAGCAAAATCCACTAGAGGCTTGAAGAGCTTGTGCTGCCACTGTGTGACC 1380
QY 1381 TCACTCACTTCTCTGATGCTTGTGAGCCCTTGCCTCAATCTGAGACCTTAATCATG 1440
Db 1381 TCACTCACTTCTCTGATGCTTGTGAGCCCTTGCCTCAATCTGAGACCTTAATCATG 1440
QY 1441 TCTGTGGGTATCACAGTGTGATGTTGTGTCCACAGCTGCTCAACAGAGACTTTT 1500
Db 1441 TCTGTGGGTATCACAGTGTGATGTTGTGTGTCCACAGCTGCTCAACAGAGACTTTT 1500
```

```
Db 1441 TCTGTGGGTATCACAGTGTGATGTTGTGTGTCCACAGCTGCTCAACAGAGACTTTT 1500
QY 1501 GTGTCCATGCTTGTGTCTAGAAAACAGACGTGGGAACTTATGTGAGCAGCATCCAC 1560
Db 1501 GTGTCCATGCTTGTGTCTAGAAAACAGACGTGGGAACTTATGTGAGCAGCATCCAC 1560
QY 1561 CAGTGAACAGGGTATTTGCTCTTCTTTCTTTCTTGAATCTTCTGCTGTGGGACACTTCA 1620
Db 1561 CAGTGAACAGGGTATTTGCTCTTCTTTCTTTCTTGAATCTTCTGCTGTGGGACACTTCA 1620
QY 1621 AGACTTGTGGCTGGAGGCTTATTAAGCAGCAGACAGTATCGAGGAAATGATCCATTA 1680
Db 1621 AGACTTGTGGCTGGAGGCTTATTAAGCAGCAGACAGTATCGAGGAAATGATCCATTA 1680
QY 1681 ACCTCCCTGTCCACATCTTTCGCCAATGGGAAATGATCTTTTCACAAAGCTCACACAC 1740
Db 1681 ACCTCCCTGTCCACATCTTTCGCCAATGGGAAATGATCTTTTCACAAAGCTCACACAC 1740
QY 1741 ATTTTCACAGAGATGCAATTTCTGAGCCCTTGGAGTTCCCACTGATTCAGAGAGAA 1800
Db 1741 ATTTTCACAGAGATGCAATTTCTGAGCCCTTGGAGTTCCCACTGATTCAGAGAGAA 1800
QY 1801 GTGGAAACAAGTTGATGCCCTACTATGAGCTTGACCATCAGAGCTATCGGTAATCAGA 1860
Db 1801 GTGGAAACAAGTTGATGCCCTACTATGAGCTTGACCATCAGAGCTATCGGTAATCAGA 1860
QY 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTTTAAGTTACACAGGGTCCCTGGGC 1920
Db 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTTTAAGTTACACAGGGTCCCTGGGC 1920
QY 1921 TGCATTTGAATACACTTCCCTCTGATTTTCCCTACATGAGAGACTTTTGAACCTGTG 1980
Db 1921 TGCATTTGAATACACTTCCCTCTGATTTTCCCTACATGAGAGACTTTTGAACCTGTG 1980
QY 1981 AAGCTGCCATCTGTTAATACTAAATTCCTAATAAGAGAAAAA 2032
Db 1981 AAGCTGCCATCTGTTAATACTAAATTCCTAATAAGAGAAAAA 2032

RESULT 2
AAK94229
ID AAK94229 standard; cDNA; 1979 BP.
XX
AC AAK94229;
XX
DI 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2816.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT P-PSDB; AAK93309.
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
```

XX Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.  
PS  
XX The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3' ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the 0190-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.  
CC  
XX  
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;  
Query Match 88.7%; Score 1802; DB 22; Length 1979;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1902; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 114 ACTTACAGCTTCTGCGAGCGAGTCTTTCACAGCCGCTTTCAGAGCTTCCACT 173  
DB 1 ACTTTCAGAGCTTCTGCGAGCGAGTCTTTCACAGCCGCTTTCAGAGCTTCCACT 60  
QY 174 TCAGCACAATGCTACTGCTTAAAAAATGAGAGTCTGTTTCTGTTCCAGATG 233  
DB 61 TCAGCACAATGCTACTGCTTAAAAAATGAGAGTCTGTTTCTGTTCCAGATG 120  
QY 234 CCATCTGGCTCTATCTTCCACATGTACAGCACAACATACCTCCCTGTTCTATGAGG 293  
DB 121 CCATCTGGCTCTATCTTCCACATGTACAGCACAACATACCTCCCTGTTCTATGAGG 180  
QY 294 CACAGCCGAGCGATGACAGTGTGTTGTTGTTCTGTTGCTGCTGCTGCTGCTTTT 353  
DB 181 CACAGCCGAGCGATGACAGTGTGTTGTTGTTCTGTTGCTGCTGCTGCTGCTTTT 240  
QY 354 TGGGCGAGCTTTTGGCGACAGCCAGATGTTTCTACATGATGAGCCGCTGCGACG 413  
DB 241 TGGGCGAGCTTTTGGCGACAGCCAGATGTTTCTACATGATGAGCCGCTGCGACG 300  
QY 414 TGTGATGACCTTAACAGAGCAGCCGCTGATGCTGACATAGGCTGCGGATCTGA 473  
DB 301 TGTGATGACCTTAACAGAGCAGCCGCTGATGCTGACATAGGCTGCGGATCTGA 360  
QY 474 TACGGGCGCTTCTTGTGACATGAGCGCTTGTGATGCTACATAGGAACCTGATCCC 533  
DB 361 TACGGGCGCTTCTTGTGACATGAGCGCTTGTGATGCTACATAGGAACCTGATCCC 420  
QY 534 GGAGACAGTCCAGCTTTTCACTGAGGAAACAGCCGCGCTGTGTGCTGACCTGCT 593  
DB 421 GGAGACAGTCCAGCTTTTCACTGAGGAAACAGCCGCGCTGTGTGCTGACCTGCT 480  
QY 594 GTGACATCATCCCAAGATGAATCATCCCGGGGCTCACTGAGGCTCTGTGACATG 653  
DB 481 GTGACATCATCCCAAGATGAATCATCCCGGGGCTCACTGAGGCTCTGTGACATG 540  
QY 654 AACAGCCCTTGAAGTGTGAGAGAGCGCTGCTTCAACAGCCAGTGTGCTCAAG 713  
DB 541 AACAGCCCTTGAAGTGTGAGAGAGCGCTGCTTCAACAGCCAGTGTGCTCAAG 600  
QY 714 AGTGGCCTTCTTCAACCTGACATCCCTTACCCGCTGCTGGAAGAGCCCTCCCTAAC 773  
DB 601 AGTGGCCTTCTTCAACCTGACATCCCTTACCCGCTGCTGGAAGAGCCCTCCCTAAC 660  
QY 774 TGCATATCTGACAGTGTGCGGAGCCCGGGCGCTGTTCCCTCCGAGAGCGACA 833  
DB 661 TGCATATCTGACAGTGTGCGGAGCCCGGGCGCTGTTCCCTCCGAGAGCGACA 720  
QY 834 AGGAGATCTCATGATGAGAGTGCATTTGATGAGGAGCATGAGCAGAACTCAGA 893  
DB 721 AGGAGATCTCATGATGAGAGTGCATTTGATGAGGAGCATGAGCAGAACTCAGA 780

QY 894 AGAGGAGCACCCTACTATGTGATGACAGTCTGCGCAAGCCAGCTGGAGATCTCA 953  
DB 781 AGAGGAGCACCCTACTATGTGATGACAGTCTGCGCAAGCCAGCTGGAGATCTCA 840  
QY 954 AAGACATCCAGTCTTGGCCAGAGCCCTGACAGAGCGTACTGCTTGGCTATGAG 1013  
DB 841 AAGACATCCAGTCTTGGCCAGAGCCCTGACAGAGCGTACTGCTTGGCTATGAG 900  
QY 1014 ACTGAGCTGAGCCCTGAGCCAGAGCTCCGATGTATGAAATTCGTGGAGTTGAT 1073  
DB 901 ACTGAGCTGAGCCCTGAGCCAGAGCTCCGATGTATGAAATTCGTGGAGTTGAT 960  
QY 1074 TCTTGGCCATTTGACAGCTGAGTGCATATCATACCCGAGGAGGAGTGGTGGC 1133  
DB 961 TCTTGGCCATTTGACAGCTGAGTGCATATCATACCCGAGGAGGAGTGGTGGC 1020  
QY 1134 AGCTTTTCCACAAATGCGAGGATGCGCTTATGTCTCCAGGCTTGGCGTGGTCT 1193  
DB 1021 AGCTTTTCCACAAATGCGAGGATGCGCTTATGTCTCCAGGCTTGGCGTGGTCT 1080  
QY 1194 TGGCTATGAAAGTCTTCTGACCTTCAAGAAAGCTTGGGATGCTATGATTTGCTGG 1253  
DB 1081 TGGCTATGAAAGTCTTCTGACCTTCAAGAAAGCTTGGGATGCTATGATTTGCTGG 1140  
QY 1254 GCTACCGCAGCTCAGATGTGAAAGAAAGAGAAACCTGTGCTGATCTTCTGCTA 1313  
DB 1141 GCTACCGCAGCTCAGATGTGAAAGAAAGAGAAACCTGTGCTGATCTTCTGCTA 1200  
QY 1314 CCGTGAAGTCTCCCTGAGCAATTCACATGAGAGGTTGAGAGGCTTGTCTGCCACTGT 1373  
DB 1201 CCGTGAAGTCTCCCTGAGCAATTCACATGAGAGGTTGAGAGGCTTGTCTGCCACTGT 1260  
QY 1374 GTGAGCTGAGCAGCTTCTGTAAGTCTTGAAGTCTGAGAGCTTGTGAGAGCTTAA 1433  
DB 1261 GTGAGCTGAGCAGCTTCTGTAAGTCTTGAAGTCTGAGAGCTTGTGAGAGCTTAA 1320  
QY 1434 CTACATGCTGTGGGTATCACACTGATGTGATTTGTCCACAGCTGCTCAAGCAGAG 1493  
DB 1321 CTACATGCTGTGGGTATCACACTGATGTGATTTGTCCACAGCTGCTCAAGCAGAG 1380  
QY 1494 GACTTTGTGTCATGCTGCTGTAGAAAGAGAGCTGGGAGCTTTGTGAGCAGCAG 1553  
DB 1381 GACTTTGTGTCATGCTGCTGTAGAAAGAGAGCTGGGAGCTTTGTGAGCAGCAG 1440  
QY 1554 ATCCACAGTGAAGAGGATGCTCTTCTTCTTCTTCTGATCTTCTGCTGCGAG 1613  
DB 1441 ATCCACAGTGAAGAGGATGCTCTTCTTCTTCTTCTGATCTTCTGCTGCGAG 1500  
QY 1614 ACTTCAGAGACTTTGTGGCTGAGGCTTATTAACAGACAGACATATCAGTGAATGA 1673  
DB 1501 ACTTCAGAGACTTTGTGGCTGAGGCTTATTAACAGACAGACATATCAGTGAATGA 1560  
QY 1674 TCCATTAACCTCCCTGTCACATCTTGGCCAAATGAGATGCTTCCACAAAGAGCT 1733  
DB 1561 TCCATTAACCTCCCTGTCACATCTTGGCCAAATGAGATGCTTCCACAAAGAGCT 1620  
QY 1734 CACACAGATTTTCCACAGAGATGCAATTTGAGCCCTTGGAGTCCAGTGAATCAAG 1793  
DB 1621 CACACAGATTTTCCACAGAGATGCAATTTGAGCCCTTGGAGTCCAGTGAATCAAG 1680  
QY 1794 GAAGAGTGGGAGAAAGGTTGATGCTACTATGAGCTTGACATGACAGCTATGCT 1853  
DB 1681 GAAGAGTGGGAGAAAGGTTGATGCTACTATGAGCTTGACATGACAGCTATGCT 1740  
QY 1854 AATCAGAAATATGAAACAAATCTGTGACAAAGAGAGAGCTTTAAGTTACAGGCTG 1913  
DB 1741 AATCAGAAATATGAAACAAATCTGTGACAAAGAGAGAGCTTTAAGTTACAGGCTG 1800  
QY 1914 CCGTGGCTGCAATTTGAATATCTTCCCTGCTGCAATTTTCCATACATATGAAGCTTGG 1973  
DB 1801 CCGTGGCTGCAATTTGAATATCTTCCCTGCTGCAATTTTCCATACATATGAAGCTTGG 1860



```

Db 1080 CAGACCTGGGTGCATACATCACACCCAGAGCGAAGGGCATGGGTATCCACGCTTTCACACA 1139
Qy 1148 AATCCAGGAGATGCCCTTAATGTCTCCAGGCTGGCGCTGTCTTCCCTATGAAGA 1207
Db 1140 AATCCAGGAGATGCCCTTAATGTCTCCAGGCTGGCGCTGTCTTCCCTATGAAGA 1199
Qy 1208 GTTCTCGACTTCAGAAAGGCTGGGGATGCCATGAATTTGCTGGGCTACCCGCCAGTC 1267
Db 1200 GTTCTCGACTTCAGAAAGGCTGGGGATGCCATGAATTTGCTGGGCTACCCGCCAGTC 1259
Qy 1268 AGATCTAGACAGAACAGAAACCTGTGTGATCTTCTGTCTACCTGGACTGTCCCT 1327
Db 1260 AGATCTAGACAGAACAGAAACCTGTGTGATCTTCTGTCTACCTGGACTGTCCCT 1319
Qy 1328 GAGCAAAATCCACTAGAGGGTTGAGAGGCTTGTGCTGCCACCTGGTGTACGCTCAGTCA 1387
Db 1320 GAGCAAAATCCACTAGAGGGTTGAGAGGCTTGTGCTGCCACCTGGTGTACGCTCAGTCA 1379
Qy 1388 CTTCCTCTGATGATGCTTGTGAGCCCTGGCTACATCTCTGAGCCTTAACATACATCTGTGG 1447
Db 1380 CTTCCTCTGATGATGCTTGTGAGCCCTGGCTACATCTCTGAGCCTTAACATACATCTGTGG 1439
Qy 1448 GTATCACACTGAGTGTGAGTTGTGTCCACAGTGTCTCAGCAGAGAGACTTTTGTGTCCA 1507
Db 1440 GTATCACACTGAGTGTGAGTTGTGTCCACAGTGTCTCAGCAGAGAGACTTTTGTGTCCA 1499
Qy 1508 TGCTTGTGTCTAGAAACAGACTGGGGACCTTATGTGACAGACACATCCACAGAGTGA 1567
Db 1500 TGCTTGTGTCTAGAAACAGACTGGGGACCTTATGTGACAGACACATCCACAGAGTGA 1559
Qy 1568 ACAGGGATATGGCTCTCTCTTCTTCTGATCTCTCTGCTGTCTGCGGACAGACTTCAGAGACTTT 1627
Db 1560 ACAGGGATATGGCTCTCTCTTCTTCTTCTGATCTCTCTGCTGTCTGCGGACAGACTTCAGAGACTTT 1619
Qy 1628 GTGGCCTGGAGGCTATTAGACACAGACAGATCAATGGAATTTGATCCATTAACCTCCC 1687
Db 1620 GTGGCCTGGAGGCTATTAGACACAGACAGATCAATGGAATTTGATCCATTAACCTCCC 1679
Qy 1688 TGTCCACATCTTGGCCCAATGGGATGATCTTTCACCAAGAGCTCACAGAGATTTTCC 1747
Db 1680 TGTCCACATCTTGGCCCAATGGGATGATCTTTCACCAAGAGCTCACAGAGATTTTCC 1739
Qy 1748 ACAGAGATGCAAAATTCGAGGCTTGGAGTTCCAGTGTGATTCAGAGAGAGAGAGTGGAA 1807
Db 1740 ACAGAGATGCAAAATTCGAGGCTTGGAGTTCCAGTGTGATTCAGAGAGAGAGAGTGGAA 1799
Qy 1808 CAAGGTGGATGGCTCTTATGAGCTTGAACATCAGCATGATGCTATGATGAGAAATATGA 1867
Db 1800 CAAGGTGGATGGCTCTTATGAGCTTGAACATCAGCATGATGCTATGATGAGAAATATGA 1859
Qy 1868 AACAAATCTCTGACAAAAGAGCAAGCTTAAATTACAGGGTGGCTGCTGATTT 1927
Db 1860 AACAAATCTCTGACAAAAGAGCAAGCTTAAATTACAGGGTGGCTGCTGATTT 1919
Qy 1928 GAATATACACTCCCTCTGATTTTCCATACATAGAAACATTTGACCTGTAAGCTGC 1987
Db 1920 GAATATACACTCCCTCTGATTTTCCATACATAGAAACATTTGACCTGTAAGCTGC 1979
Qy 1988 CATCTGTTAATCTAATTAATCCCAATATAG 2017
Db 1980 CATCTGTTAATCTAATTAATCCCAATATAG 2009

```

```

XX Human; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
KW L-selectin; L-selectin sulfoltransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
KW antileukemia; antiporiatic; antidiabetic; dermatological;
KW antiallergic.
XX Homo sapiens.
XX OS
XX Key Location/Qualifiers
XX CDS 111..1253
XX FT /*tag= a
XX FT /product= "Human LST-2"
XX WO200185177-A1.
XX PD 15-NOV-2001.
XX PF 10-MAY-2001; 2001WO-US15452.
XX PR 11-MAY-2000; 2000US-0569320.
XX PA (BURN-) BURNHAM INST.
XX PI Fukuda M, Yeh J, Hiraoka N;
XX DR MPI: 2002-075226/10.
XX DR P-PSDB: MAU11274.
XX PS New enzyme, useful for modifying acceptor molecule, comprises an
PT isolated L-selectin sulfoltransferase-2 that directs expression of
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
PT intestinal GlcNAc 6-sulfoltransferase
XX Claim 19; Fig 4; 98pp; English.
XX CC The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated
CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
CC invention also provides a method of treating or preventing an
CC L-selectin-mediated condition by reducing the expression or activity of a
CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3gnt,
CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
CC sulfoltransferase-2 (LST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents cDNA encoding human LST-2.
XX SQ Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other:
XX
XX Query Match 59.6%; Score 1211; DB 24; Length 1333;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 161 AAGGTCTCCACTGACGACCAATGCTACTGCTAATAAAATGAGCTCTGCTGTTTG 220
Db 72 AAGGTCTCCACTGACGACCAATGCTACTGCTAATAAAATGAGCTCTGCTGTTTG 131
Qy 221 GTTCCCAATGAGGCAATGCTGCTATTTCTCCACATGATACAGCCACATCAGCTCC 280
Db 132 GTTCCCAATGAGGCAATGCTGCTATTTCTCCACATGATACAGCCACATCAGCTCC 191

```

```
QY 281 CTGTCTATGAGGACACAGCCCGAGCGCATGCACTGCTGCTTCTCTCTCTGCGGCTCT 340
    |||||||
Db 192 CTGTCTATGAGGACACAGCCCGAGCGCATGCACTGCTGCTGCTCTCTCTCTGCGGCTCT 251
QY 341 GGCCTCTCTTTTGTGGGACAGCTTTTGGGACGACCCAGATGTTTCTACTGATGAG 400
    |||||||
Db 252 GGCCTCTCTTTTGTGGGACAGCTTTTGGGACGACCCAGATGTTTCTACTGATGAG 311
QY 401 CCCGCTGAGCAGCTGTGATACCTTCAAGCAGAGCAGCCGCTTGATGCTGACATGGCT 460
    |||||||
Db 312 CCCGCTGAGCAGCTGTGATACCTTCAAGCAGAGCAGCCGCTTGATGCTGACATGGCT 371
QY 461 GTGGGGATCTGATACGAGGCGCTCTTGTGACACATGAGCGCTTGTGATGCTTACATG 520
    |||||||
Db 372 GTGGGGATCTGATACGAGGCGCTCTTGTGACACATGAGCGCTTGTGATGCTTACATG 431
QY 521 GAACCTGCTCCCCGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCCGGGCGCTGTGT 580
    |||||||
Db 432 GAACCTGCTCCCCGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCCGGGCGCTGTGT 491
QY 581 TCTGCACCTGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGACAG 640
    |||||||
Db 492 TCTGCACCTGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGACAG 551
QY 641 CTCTGTGACAGTCAACAGCCCTTGTAGGTGTGAGAGAGGCTGCGCTCTCTACAGCCAC 700
    |||||||
Db 552 CTCTGTGACAGTCAACAGCCCTTGTAGGTGTGAGAGAGGCTGCGCTCTCTACAGCCAC 611
QY 701 GTGCTGCTCAAGAGAGGCGCTTCTTCAACCTGCACTGCTCTACCCGCTGCTGAAGAC 760
    |||||||
Db 612 GTGCTGCTCAAGAGAGGCGCTTCTTCAACCTGCACTGCTCTACCCGCTGCTGAAGAC 671
QY 761 CCCTCCCTCAACCTGCACTATCTGTCACCTGCTCCGAGACCCCGGCGCTGCTCTTCC 820
    |||||||
Db 672 CCCTCCCTCAACCTGCACTATCTGTCACCTGCTCCGAGACCCCGGCGCTGCTCTTCC 731
QY 821 CGAAGAGGACAAAGGAGATCTCTATGATGACAGTGCATTTGATGGGAGACATGAG 880
    |||||||
Db 732 CGAAGAGGACAAAGGAGATCTCTATGATGACAGTGCATTTGATGGGAGACATGAG 791
QY 881 CAGAACTCAAGAGAGAGACCAACCTTACTATGTGATGACAGTGCATCTCCAAAGCAG 940
    |||||||
Db 792 CAAAACTCAAGAGAGAGACCAACCTTACTATGTGATGACAGTGCATCTCCAAAGCAG 851
QY 941 CTGAGATCTCAAGACCATCATCTCTTGGCCAGGCGCTGACAGAGCCTTACTGCTT 1000
    |||||||
Db 852 CTGAGATCTCAAGACCATCATCTCTTGGCCAGGCGCTGACAGAGCCTTACTGCTT 911
QY 1001 GTGGGCTATGAGAGCTGGCTGCGAGCCCTGTGGCCAGAGCTTCCCAATGTATGATTC 1060
    |||||||
Db 912 GTGGGCTATGAGAGCTGGCTGCGAGCCCTGTGGCCAGAGCTTCCCAATGTATGATTC 971
QY 1061 GTGGGATGCAATCTTCCCATCTTACAGACCTGGGTGATACATCACCCGAGGCAAG 1120
    |||||||
Db 972 GTGGGATGCAATCTTCCCATCTTACAGACCTGGGTGATACATCACCCGAGGCAAG 1031
QY 1121 GGCATGGGTGACCAAGCTTTTCCACAAATGCCAGGATGCCCTTATGTCTCCAGGCT 1180
    |||||||
Db 1032 GGCATGGGTGACCAAGCTTTTCCACAAATGCCAGGATGCCCTTATGTCTCCAGGCT 1091
QY 1181 TGGGCTGTGCTTTTCCCTATGAAAAGTTTCTGACTTGAAGAAGCCTTGGGATGCC 1240
    |||||||
Db 1092 TGGGCTGTGCTTTTCCCTATGAAAAGTTTCTGACTTGAAGAAGCCTTGGGATGCC 1151
QY 1241 ATGATTTTGTGGGCTACCCGACAGTCAAGTCTGAACAAAGAGAAACCTGTTGCTG 1300
    |||||||
Db 1152 ATGATTTTGTGGGCTACCCGACAGTCAAGTCTGAACAAAGAGAAACCTGTTGCTG 1211
QY 1301 GATTTTGTCTTACTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTTGAAGGCTTT 1360
    |||||||
Db 1212 GATTTTGTCTTACTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTTGAAGGCTTT 1271
QY 1361 GCTGCCACCTGGGTCTCAGGCTCAGTCACTTCTCTGATGCTTGAAGCTTGGCTTACAT 1420
```

```
Db 1272 GGTGCCACCTGGGTGACGCTCAGTCACTTCTCTGAATGCTTGTAGGCTTGCTTACAT 1331
    |||||||
QY 1421 CT 1422
    ||
Db 1332 CT 1333

RESULT 5
AAK91803
ID AAK91803 standard; cDNA: 877 BP.
AC AAK91803;
XX
DT 06-NOV-2001 (first entry)
DE Human cDNA 5'-end sequence, SEQ ID NO: 263.
DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP113094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI: 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 36.5%; Score 741; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 1,1e-305;
Matches 741; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ACTTTCACAGCTTCCCTGGAGCGAGTCTTTCAGAGCCGCTTGCAGAGTCTCCACT 173
    |||||||
Db 1 ACTTTCACAGCTTCCCTGGAGCGAGTCTTTCAGAGCCGCTTGCAGAGTCTCCACT 60
QY 174 TCAGCAATGCTACTGCGCTTAAATAAATGAAGTCTGCTGTGCTTGGCTTCCAGATG 233
    |||||||
Db 61 TCAGCAATGCTACTGCTTAAATAAATGAAGTCTGCTGTGCTTGGCTTCCAGATG 120
QY 234 CCATCTGGCTCTATTCTTCACATGTACAGCCACAAATCAGTCCCTGTCTATGAGG 293
    |||||||
```

DB 121 CCATCTTGCTCTATCTTCCACATGTACACCAACATCAGTCCCTCTCTATGAAG 180  
QY 294 CACAGCCCGAGGCGATGCACGTGTGTCTCTCTCCGCGCTCTGAGCTCTCTTTG 353  
DB 181 CACAGCCCGAGGCGATGCACGTGTGTCTCTCTCCGCGCTCTGAGCTCTCTTTG 240  
QY 354 TGGGCGAGCTTTTGGCGAGCACCGAGATGTTTCTACCTGATGAGACCCGCTGACAG 413  
DB 241 TGGGCGAGCTTTTGGCGAGCACCGAGATGTTTCTACCTGATGAGACCCGCTGACAG 300  
QY 414 TGTGATGACCTTCAAGCAGACGCGGCTGATGATGACATGGGTGGCGGATCTGA 473  
DB 301 TGTGATGACCTTCAAGCAGACGCGGCTGATGATGACATGGGTGGCGGATCTGA 360  
QY 474 TACGGGCGCTTCTTGTGCGAATGAGCGCTTTGATGCTTACATGAACTGTGCCCC 533  
DB 361 TACGGGCGCTTCTTGTGCGAATGAGCGCTTTGATGCTTACATGAACTGTGCCCC 420  
QY 534 GAGAGACGTCCAGCTCTTCTTCACTGGAGAAACAGCGCGGCGCTGCTGACCTGCT 593  
DB 421 GAGAGACGTCCAGCTCTTCTTCACTGGAGAAACAGCGCGGCGCTGCTGACCTGCT 480  
QY 594 GTGACATCATCCCAAGATGAATCATCCCGGCGCTCACTGCAAGCTCTGTGACATC 653  
DB 481 GTGACATCATCCCAAGATGAATCATCCCGGCGCTCACTGCAAGCTCTGTGACATC 540  
QY 654 AACAGCCCTTTGAGT 713  
DB 541 AACAGCCCTTTGAGT 600  
QY 714 AGGAGCGCTTCTTCAAGCTGCACTGCTCTTCAAGCTGCTGCTGCTGCTGCTGCT 773  
DB 601 AGGAGCGCTTCTTCAAGCTGCACTGCTCTTCAAGCTGCTGCTGCTGCTGCTGCT 660  
QY 774 TGCATATGCTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 833  
DB 661 TGCATATGCTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720  
QY 834 AGGAGATCTCATGATGTGACA 854  
DB 721 AGGAGATCTCATGATGTGACA 741

RESULT 6  
AAK93921  
ID AAK93921 standard; cDNA; 877 BP.  
XX AAK93921;  
AC AAK93921;  
XX 06-NOV-2001 (first entry)  
DE Human cDNA clone representative sequence, SEQ ID NO: 2381.  
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
KM Homo sapiens.  
OS Homo sapiens.  
XX EP1130094-A2.  
PN EP1130094-A2.  
PD 05-SEP-2001.  
XX 07-JUL-2000; 2000EP-0114089.  
PF 08-JUL-1999; 99JP-0194486.  
XX 11-JAN-2000; 2000JP-0118774.  
PR 02-MAY-2000; 2000JP-0183765.  
XX (HELI-) HELIX RES INST.  
PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX WPI; 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -  
XX  
PS  
XX Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence was used as the  
CC representative sequence to identify the clone.  
CC homology searches to identify the clone.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
XX  
XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;  
Query Match 36.5%; Score 741; DB 22; Length 877;  
Best Local Similarity 100.0%; Pred. No. 1.1e-305;  
Matches 741; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 114 ACTTTCACAGCTTCCCTGGAGCGAGTGTCTTCTCAAGCCGCTTGAAGGCTTCCACT 173  
DB 1 ACTTTCACAGCTTCCCTGGAGCGAGTGTCTTCTCAAGCCGCTTGAAGGCTTCCACT 60  
QY 174 TCAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGCTTGTGTTTCCAGATG 233  
DB 61 TCAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGCTTGTGTTTCCAGATG 120  
QY 234 CCATCTTGCTCTATCTTCCACATGTACAGCCACATAGTCCCTGTCTGTAAG 293  
DB 121 CCATCTTGCTCTATCTTCCACATGTACAGCCACATAGTCCCTGTCTGTAAG 180  
QY 294 CACAGCCCGAGGCGATGCACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 353  
DB 181 CACAGCCCGAGGCGATGCACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240  
QY 354 TGGGCGAGCTTTTGGCGAGCACCGAGATGTTTCTACCTGATGAGACCCGCTGACAG 413  
DB 241 TGGGCGAGCTTTTGGCGAGCACCGAGATGTTTCTACCTGATGAGACCCGCTGACAG 300  
QY 414 TGTGATGACCTTCAAGCAGACGCGGCTGATGATGACATGGGTGGCGGATCTGA 473  
DB 301 TGTGATGACCTTCAAGCAGACGCGGCTGATGATGACATGGGTGGCGGATCTGA 360  
QY 474 TACGGGCGCTTCTTGTGCGAATGAGCGCTTTGATGCTTACATGAACTGTGCCCC 533  
DB 361 TACGGGCGCTTCTTGTGCGAATGAGCGCTTTGATGCTTACATGAACTGTGCCCC 420  
QY 534 GAGAGACGTCCAGCTCTTCTTCACTGGAGAAACAGCGCGGCGCTGCTGACCTGCT 593  
DB 421 GAGAGACGTCCAGCTCTTCTTCACTGGAGAAACAGCGCGGCGCTGCTGACCTGCT 480  
QY 594 GTGACATCATCCCAAGATGAATCATCCCGGCGCTCACTGCAAGCTCTGTGACATC 653  
DB 481 GTGACATCATCCCAAGATGAATCATCCCGGCGCTCACTGCAAGCTCTGTGACATC 540  
QY 654 AACAGCCCTTTGAGT 713  
DB 541 AACAGCCCTTTGAGT 600  
QY 714 AGGAGCGCTTCTTCAAGCTGCACTGCTCTTCAAGCTGCTGCTGCTGCTGCTGCT 773  
DB 601 AGGAGCGCTTCTTCAAGCTGCACTGCTCTTCAAGCTGCTGCTGCTGCTGCTGCT 660  
QY 774 TGCATATGCTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 833  
DB 661 TGCATATGCTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720



QY 834 AGGAGATCTCATGATTGACA 854  
|||||  
Db 721 AGGAGATCTCATGATTGACA 741

## RESULT 7

ABK547724  
ID ABK54724 standard; cDNA; 517 BP.

AC ABK54724:

DT 18-JUN-2002 (first entry)

XX Human colon cancer-associated cDNA, SEQ ID NO 194.

XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

XX Homo sapiens.

XX W0200212280-A2.

XX 14-FEB-2002.

XX 30-JUL-2001; 2001WO-US23826.

XX 03-AUG-2000; 2000US-223265P.

XX 02-OCT-2000; 2000US-237406P.

XX 20-MAR-2001; 2001US-277495P.

XX 03-JUL-2001; 2001US-302702P.

XX (CORI-) CORIXA CORP.

XX Pyle RA, Xu J, Secrist H;

XX WPI; 2002-257462/30.

XX Novel polynucleotide encoding colon tumour polypeptides, useful as

XX vaccines for treating colon cancers -

XX Claim 1; Page 206; 425pp; English.

XX The invention relates to isolated polynucleotides (i) encoding colon

XX tumour polypeptides (ii). (i) is useful for stimulating an immune

XX response in a patient and treating colon cancer in a patient.

XX Oligonucleotides derived from (i) are useful for determining the presence

XX of cancer in a patient. (i) and (ii) are useful in pharmaceutical

XX compositions, e.g. vaccines, and other compositions for the diagnosis

XX and treatment of colon cancer. A composition comprising a first component

XX selected from physiologically acceptable carriers and immunostimulants,

XX and an antigen-presenting cell expressing (ii) is useful for inhibiting

XX development of cancer in a patient. (i) is useful in the design and

XX preparation of ribozyme molecules for inhibiting expression of tumour

XX polypeptides and (i). ABK54531-ABK5464 represent human colon cancer cDNA

XX sequences of the invention.

XX Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

XX Query Match 24.9%; Score 505; DB 24; Length 517;

XX Best Local Similarity 100.0%; Pred. No. 3.8e-205;

XX Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1355 GCGTTTGCTGCCACCTGGTGCACGCTCAGTCACTTCTCGAATGCTTCTGAGCCTTGC 1414  
|||||  
Db 13 GCGTTTGCTGCCACCTGGTGCACGCTCAGTCACTTCTCGAATGCTTCTGAGCCTTGC 72  
1415 CTACATCTCTAGCCTTAAGTACATGCTGTGCGTATCACACTGAGTGTGAGTGTGCC 1474  
|||||  
Db 73 CTACATCTCTAGCCTTAAGTACATGCTGTGCGTATCACACTGAGTGTGAGTGTGCC 132  
1475 ACAGTGTCTCAAGAGAAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAAAACAGACTGGGG 1534  
|||||  
Db 133 ACAGTGTCTCAAGAGAAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAAAACAGACTGGGG 192

QY 1535 AACCTTATGTGAGCAGACACATCCACAGTGAACAGGATATTGCTCTTCTTTTCTT 1594  
|||||  
Db 193 AACCTTATGTGAGCAGACACATCCACAGTGAACAGGATATTGCTCTTCTTTTCTT 252  
QY 1595 GATCTTCCTGCTGGGAGACTTTCAGAGACTTTGGGCGCTGGAGGCTTCTTAAGCAGAC 1654  
|||||  
Db 253 GATCTTCCTGCTGGGAGACTTTCAGAGACTTTGGGCGCTGGAGGCTTCTTAAGCAGAC 312  
QY 1655 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCCACATCTTGCCCAATGGGGAATG 1714  
|||||  
Db 313 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCCACATCTTGCCCAATGGGGAATG 372  
QY 1715 GATCTTCACCAAGAGCTTCACGAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGG 1774  
|||||  
Db 373 GATCTTCACCAAGAGCTTCACGAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGG 432  
QY 1775 AGTTCACAGTGAATTCAGAGAGAGTGGGAACAAGTTGATGCTTACTATGAGCTT 1834  
|||||  
Db 433 AGTTCACAGTGAATTCAGAGAGAGTGGGAACAAGTTGATGCTTACTATGAGCTT 492  
QY 1835 GACCATCACAGCTTATCGGTATCAG 1859  
|||||  
Db 493 GACCATCACAGCTTATCGGTATCAG 517

## RESULT 8

AAK92588/C  
ID AAK92588 standard; cDNA; 548 BP.

XX AAK92588;

XX 06-NOV-2001 (first entry)

XX Human cDNA 3'-end sequence, SEQ ID NO: 1048.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their

XX use in genetic manipulation -

XX Claim 3; SEQ ID NO 1048; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA

XX clones. 830 cDNA molecules encoding a human protein have been

XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

XX molecules have been determined. Primers for synthesizing the full length

XX cDNA are useful for clarifying the function of the protein encoded by

XX the cDNA. The full length clones were obtained by construction of full

XX length enriched cDNA libraries that were synthesized by the oligo-capping

XX method. The primers enable the production of the full length cDNA easily

XX without any special methods. The present sequence is the nucleotide

XX sequence of the 3'-end of a cDNA provided in the invention.

XX Note: The sequence data for this patent did not form part of the printed

CC Specification, but was obtained in CD-ROM format directly from EPO.  
XX  
SQ Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;  
Query Match 17.3%; Score 351; DB 22; Length 548;  
Best Local Similarity 100.0%; Pred. No. 1.5e-139;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1667 GAATTGATTCATAAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTTCACCA 1726  
DB 426 GAATTGATTCATAAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTTCACCA 367  
QY 1727 AAGAGCTCACCAGATTTTCCACAGATGCAATCTTGAGCCCTTGAGATTTCCAGTGG 1786  
DB 366 AAGAGCTCACCAGATTTTCCACAGATGCAATCTTGAGCCCTTGAGATTTCCAGTGG 307  
QY 1787 ATTCAGAGGAAGAGTGGAGACAGGTTGGATGCTACTATGAGCTTTGACATCAGC 1846  
DB 306 ATTCAGAGGAAGAGTGGAGACAGGTTGGATGCTACTATGAGCTTTGACATCAGC 247  
QY 1847 TATCGGTATTCAGAAATATGACAAATATCTGACAAAGAGCAAGCTTTAAGTTCA 1906  
DB 246 TATCGGTATTCAGAAATATGACAAATATCTGACAAAGAGCAAGCTTTAAGTTCA 187  
QY 1907 CAGGCTCCTGGGCTGCATTTGATATCACTTCCCTCTGCATTTCCATCAGATGAA 1966  
DB 186 CAGGCTCCTGGGCTGCATTTGATATCACTTCCCTCTGCATTTCCATCAGATGAA 127  
QY 1967 GACTTTGACCTGTGAGACCTGCTGTGTATACATAAATCCCAATTAAG 2017  
DB 126 GACTTTGACCTGTGAGACCTGCTGTGTATACATAAATCCCAATTAAG 76  
RESULT 9  
AAC76156  
ID AAC76156 standard; cDNA; 2988 BP.  
XX  
AC AAC76156;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.  
XX  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
XX vlnervary; antiporiatic; antiparkinsonian; noctropic; neuroprotective;  
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;  
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;  
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
XX cholesterol ester storage; systemic lupus erythematosus; infection;  
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;  
XX thrombosis; contraceptive; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PE 31-MAR-2000; 2000MO-US08621.  
XX  
PR 31-MAR-1999; 9905-0127607.  
XX 02-APR-1999; 9905-0127636.  
XX 05-APR-1999; 9905-0127728.  
XX 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.

PI Shinkels RA, Leach M;  
XX  
XX WPI: 2000-602362/57.  
DR P-PSDB: AAB41947.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
XX useful for treating e.g. cancers, proliferative disorders,  
XX neurodegenerative disorders and cardiovascular disease -  
Claim 5; Page 2597-2599; 5507pp; English.  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vlnervary;  
CC antiporiatic; antiparkinsonian; noctropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and anaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;  
Query Match 15.1%; Score 306; DB 21; Length 2988;  
Best Local Similarity 100.0%; Pred. No. 2e-120;  
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 506 TTGTAGTCTACATGAGAACTGCTCCCGGAGACAGCTCCAGCTTTCAGTGGAGAC 565  
DB 1480 TTGTAGTCTACATGAGAACTGCTCCCGGAGACAGCTCCAGCTTTCAGTGGAGAC 1539  
QY 566 AGCCGGGCCCTGTGTTGACACCTGCTGTGACATCATCCAGAGATGAATCATCCCC 625  
DB 1540 AGCCGGGCCCTGTGTTGACACCTGCTGTGACATCATCCAGAGATGAATCATCCCC 1599  
QY 626 CGGCTCACTGAGGCTCCTGTGACGACGACGCTTGAAGGAGTGAAGAAGGCTGC 685  
DB 1600 CGGCTCACTGAGGCTCCTGTGACGACGACGCTTGAAGGAGTGAAGAAGGCTGC 1659  
QY 686 CGCTCTACAGCAGCAGTGTGCTCAAGAGAGTGCCTTCTCAACCTGACAGTCCCTTAC 745  
DB 1660 CGCTCTACAGCAGCAGTGTGCTCAAGAGAGTGCCTTCTCAACCTGACAGTCCCTTAC 1719  
QY 746 CCGCTCTGAAGAGCCCTCCTCAACCTGATATCGGACACTGTGCGGAGACCCCGG 805  
DB 1720 CCGCTCTGAAGAGCCCTCCTCAACCTGATATCGGACACTGTGCGGAGACCCCGG 1779  
QY 806 GCGGTG 811  
DB 1780 GCGGTG 1785  
RESULT 10  
ABK54794  
ID ABK54794 standard; cDNA; 389 BP.  
XX  
AC ABK54794;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human colon cancer-associated cDNA, SEQ ID NO 264.

XX Human: colon cancer; immunogenic; vaccine; tumour; gene; ss.  
KM Homo sapiens.  
OS  
XX W0200212280-A2.  
PN  
XX 14-FEB-2002.  
PD  
XX 30-JUL-2001; 2001WO-US23826.  
PF  
XX 03-AUG-2000; 2000US-223265P.  
PR 02-OCT-2000; 2000US-237406P.  
PR 20-MAR-2001; 2001US-227495P.  
PR 03-JUL-2001; 2001US-302702P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Pyle RA, Xu J, Secrist H;  
XX WPI: 2002-257462/30.  
XX  
XX Novel polynucleotide encoding colon tumour polypeptides, useful as  
PT vaccines for treating colon cancers  
XX  
XX Claim 1; Page 225; 425pp; English.  
XX  
XX The invention relates to isolated polynucleotides (I) encoding colon  
CC tumour polypeptides (II). (I) is useful for stimulating an immune  
CC response in a patient and treating colon cancer in a patient.  
CC Oligonucleotides derived from (I) are useful for determining the presence  
CC of cancer in a patient. (I) and (II) are useful in pharmaceutical  
CC compositions, e.g. vaccines, and other compositions comprising a first component  
CC and treatment of colon cancer. A composition comprising a first component  
CC selected from physiologically acceptable carriers and immunostimulants,  
CC and an antigen-presenting cell expressing (II) is useful for inhibiting  
CC development of cancer in a patient. (I) is useful in the design and  
CC preparation of ribozyme molecules for inhibiting expression of tumour  
CC polypeptides and (I). ABK54531-ABK5464 represent human colon cancer cDNA  
CC sequences of the invention.  
XX  
XX Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;  
SQ  
Query Match 14.1%; Score 286; DB 24; Length 389;  
Best Local Similarity 100.0%; Pred. No. 7.5e-112;  
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1732 CTCACGACATTTCCAGAGAGATGCAATCTGAGCCTTGAGCTCCCGTGGATTCA 1791  
DB 13 CTCACGACATTTCCAGAGAGATGCAATCTGAGCCTTGAGCTCCCGTGGATTCA 72  
XX  
XX 1792 AGAAGAGAGTGGAGACAGGTTGGATGCTTACTTATGAGCTTGACCATCAGACTATCG 1851  
DB 73 AGAAGAGAGTGGAGACAGGTTGGATGCTTACTTATGAGCTTGACCATCAGACTATCG 132  
XX  
XX 1852 GTAATCAGAAATATGAACAGAAATCTGCGACAAAGAGCAGCTTTAACTTACAGGG 1911  
DB 133 GTAATCAGAAATATGAACAGAAATCTGCGACAAAGAGCAGCTTTAACTTACAGGG 192  
XX  
XX 1912 TGCCTGGGCTCATTTGATATCATCTCCCTGCACTTTCCCATACATAGAGACTT 1971  
DB 193 TGCCTGGGCTCATTTGATATCATCTCCCTGCACTTTCCCATACATAGAGACTT 252  
XX  
XX 1972 TGACCTGGAAGCTGCCATCTGTTATATCTAAATTCCTCAATAG 2017  
DB 253 TGACCTGGAAGCTGCCATCTGTTATATCTAAATTCCTCAATAG 298

XX 15-JUL-2002 (first entry)  
DT  
XX Human spliced transcript detection oligonucleotide SEQ ID NO:10822.  
DE  
XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
KM splice variant; transcriptome; oligonucleotide library; ss.  
XX  
XX Homo sapiens.  
OS  
XX W0200210449-A2.  
PN  
XX 07-FEB-2002.  
PD  
XX 20-JUL-2001; 2001WO-IB01903.  
PF  
XX 28-JUL-2000; 2000US-221607P.  
PR 02-MAY-2001; 2001US-287724P.  
PR  
XX (COMP-) COMPUGEN INC.  
XX  
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
XX WPI: 2002-257383/30.  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes  
XX  
XX Example 1; SEQ ID 10822; 47pp; English.  
XX  
XX The present invention describes oligonucleotide libraries for detecting  
CC messenger RNAs that populate a (sub-)transcriptome, where the  
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
CC transcription units that populate a genome. The library comprises  
CC several oligonucleotides, each capable of hybridizing selectively to a  
CC set of messenger RNAs transcribed from a given transcription unit of  
CC the genome, which encodes one or more messenger RNA splice variants.  
CC The oligonucleotide libraries are useful for detecting mRNAs from a  
CC biological sample, in expression profiling studies, in qualitatively or  
CC quantitatively characterizing the corresponding transcriptome, and in  
CC detecting RNA transcripts and splice variants of human or animal  
CC transcriptomes. The libraries may also be used as specialised mini  
CC libraries to detect transcripts of a sub-transcriptome under a  
CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN27253 to ABN59589 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 60 BP; 13 A; 20 C; 9 G; 18 T; 0 other;  
SQ  
Query Match 3.0%; Score 60; DB 24; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.5e-15;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1934 CACTCCCTCTGCACTTTTCCCATGACATGAGAGACTTTACCTGGAAGCTCCGATCG 1993  
DB 1 CACTCCCTCTGCACTTTTCCCATGACATGAGAGACTTTGACCTGGAAGCTCCGATCG 60

RESULT 11  
ABN38074  
ID ABN38074 standard; DNA; 60 BP.  
XX  
XX AC ABN38074;

RESULT 12  
AAD24670  
ID AAD24670 standard; cDNA; 1647 BP.  
XX  
XX AC AAD24670;



XX  
PI Rosen SD, Lee JK, Hemmerich S;  
XX  
DR WPI: 2001-138471/14.  
DR P-PSDB: AAT72640.  
XX  
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for  
PT diagnostic and therapeutic agent screening applications  
XX  
XX Claim 6; Fig 4A; 128pp; English.  
XX  
CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4  
CC beta) cDNA. GST-4 gene is found on chromosome 16q23.1.  
CC GST is a type 2 membrane protein useful for inhibiting a binding event  
CC between a selectin and a selectin ligand, which comprises contacting the  
CC selectin with a non-sulphated selectin ligand, GST and a small molecular  
CC agent that inhibits the sulphation activity of GST. GST is also useful  
CC in inhibiting a selectin mediated binding event. GST is useful in gene  
CC therapy to treat disorders such as acute or chronic inflammation,  
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis  
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's  
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious  
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,  
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
CC during transplantation.  
CC  
SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;  
Query March 2.9%; Score 59; DB 22; Length 1694;  
Best Local Similarity 100.0%; Pred. No. 3.3e-15;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 678 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 736  
|||||  
DB 676 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 734  
|||||  
RESULT 14  
AAD02699  
ID AAD02699 standard; cDNA: 2044 BP.  
XX  
AC AAD02699;  
XX  
DT 02-MAY-2001 (first entry)  
XX  
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.  
XX  
KW Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;  
KW therapy; selectin binding inhibitor; gene therapy; inflammation;  
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;  
KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;  
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;  
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;  
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;  
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;  
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;  
KW chromosome 16q23.1; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH CDS 218..1390  
FT /\*tag= a  
FT /product= "Human glycosyl sulfotransferase-4alpha  
FT (GST-4alpha)"  
FT /note= "CDS is specifically claimed as SEQ ID NO: 4  
FT in claim 6 (page no: 41) of the specification"  
XX  
XX W0200106015-A1.

PD 25-JAN-2001.  
XX  
XX 19-JUL-2000; 200GWO-US19741.  
PF  
XX 20-JUL-1999; 99US-0144694.  
PR 13-JUL-2000; 2000US-0593828.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Rosen SD, Lee JK, Hemmerich S;  
XX  
XX WPI: 2001-138471/14.  
XX P-PSDB: AAT72639.  
XX  
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for  
PT diagnostic and therapeutic agent screening applications  
XX  
XX Claim 6; Fig 1; 128pp; English.  
XX  
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4  
CC alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.  
CC GST is a type 2 membrane protein useful for inhibiting a binding event  
CC between a selectin and a selectin ligand, which comprises contacting the  
CC selectin with a non-sulphated selectin ligand, GST and a small molecular  
CC agent that inhibits the sulphation activity of GST. GST is also useful  
CC in inhibiting a selectin mediated binding event. GST is useful in gene  
CC therapy to treat disorders such as acute or chronic inflammation,  
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis  
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's  
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious  
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,  
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
CC during transplantation.  
CC Note: The present sequence is also shown in sequence listing (page  
CC no: 56) but lacks four nucleotides at its 3' end.  
CC  
SQ Sequence 2044 BP; 422 A; 645 C; 570 G; 407 T; 0 other;  
Query March 2.9%; Score 59; DB 22; Length 2044;  
Best Local Similarity 100.0%; Pred. No. 3.3e-15;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 678 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 736  
|||||  
DB 708 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 766  
|||||  
RESULT 15  
AAD02698  
ID AAD02698 standard; cDNA: 2170 BP.  
XX  
AC AAD02698;  
XX  
DT 02-MAY-2001 (first entry)  
XX  
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.  
XX  
KW Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;  
KW therapy; selectin binding inhibitor; gene therapy; inflammation;  
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;  
KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;  
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;  
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;  
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;  
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;  
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;  
KW chromosome 16q23.1; ss.  
XX  
OS Homo sapiens.  
XX  
XX

	Key	Location/Qualifiers
FH	5'UTR	9..343
FT		/+tag- a
FT	CDS	344..1516
FT		/+tag- b
FT		/product= "Human glycosyl sulfotransferase-4alpha
FT		(SST-4alpha)"
FT		/note= "CDS is specifically claimed as SEQ ID NO: 4
FT	3'UTR	in claim 6 (page no: 41) of the specification"
FT		1517..2134
FT		/+tag- c
XX		
XX	PN	WO200106015-A1.
XX	PD	25-JAN-2001.
XX	PP	19-JUL-2000; 2000WC-US19741.
XX	PR	20-JUL-1999; 99US-0144694.
XX	PR	13-JUL-2000; 2000US-0593828.
XX	PA	(REGC ) UNIV CALIFORNIA.
XX	PI	Rosen SD, Lee JK, Hemmerlich S;
XX	DR	WPI: 2001-138471/14.
XX	P-PSDB:	AAY72639.
XX	PT	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX	PS	diagnostic and therapeutic agent screening applications -
XX	PS	Claim 6; Page 62; 128pp; English.
CC	<p>The present sequence is human glycosyl sulfotransferase-4alpha (GST-4 alpha) cDNA. GST-4 gene is found on chromosome 16q23.1. GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation.</p>	
SQ	Sequence 2170 BP; 449 A; 602 C; 612 G; 427 T; 0 other;	
	<p>Query Match                    2.9%; Score 59; DB 22; Length 2170; Best Local Similarity      100.0%; Pred. No. 3.3e-15; Matches     59; Conservative     0; Mismatches     0; Indels     0; Gaps     0;</p>	
Db	678 AGGCGTCGCCCTCTACAGCCACGCTGAGGAGGGTCCGTTCACCTGCAG 736   834 AGGCGTCGCCCTCTACAGCCACGCTGAGGAGGAGGTCGTTCACTGCAG 892	
RESULT 16		
ABN89506		
ID	ABN89506 standard; cDNA: 2544 BP.	
XX		
AC	ABN89506;	
XX		
DT	05-SEP-2002 (first entry)	
XX		
DE	Human cornel N-acetylglucosamine-6-sulfotransferase cDNA SEQ ID NO:1.	
XX		
FW	Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;	

KM	corneal sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW	ophthalmological chromosome 16q22; gene; ss.
XX	
OS	Homo sapiens.
FH	Key Location/Qualifiers
FT	CDS 693..1880
FT	/tag= a
FT	/product= "N-acetylgucosamine-6-sulfotransferase"
PN	US2002061562-A1.
PD	23-MAY-2002.
PF	09-AUG-2001; 2001US-0927602.
PR	11-AUG-2000; 2000US-325773P.
XX	(FUKU/) FUKUDA M N.
PA	(AKAM/) AKAMA T O.
XX	
PI	Fukuda MN, Akama TO;
DR	WPI: 2002-507643/54.
DR	P-PsDB; ABB81554.
XX	
PT	New nucleic acid encoding corneal
PT	N-acetylgucosamine-6-sulfotransferase, useful for treatment,
PI	monitoring and diagnosis of macular corneal dystrophy -
XX	
PS	Claim 4; Fig 1A-D; 69pp; English.
XX	
CC	The present sequence encodes human corneal
CC	N-acetylgucosamine-6-sulfotransferase (I), which is able to catalyze
CC	sulfation of keratan sulfate (KS). Also described is a method for
CC	monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC	and detecting susceptibility to MCD. (I) is located to Chromosome 16q22,
CC	and has ophthalmological activity. (I) can be used to treat or prevent
CC	macular corneal dystrophy types I or II. (I) makes possible treatment
CC	of MCD without requiring keratoplasty or keratectomy.
XX	
SQ	Sequence 2544 BP; 460 A; 800 C; 733 G; 551 T; 0 other:
	Query Match 2.9%; Score 59; DB 24; Length 2544;
	Best Local Similarity 100.0%; Pred. No. 3.3e-15;
	Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	.678 AGCGCTGCGGCCTCAGACGCCACTGTCGTCCACAGAGAGTGCGCTTCAACCTGCAG 736
DB	1180 AGCGCTGCGGCCTCAGACGCCACTGTCGTCCACAGAGAGTGCGCTTCAACCTGCAG 1238
	RESULT 17
ID	ABN89533 standard; DNA: 48436 BP.
AC	ABN89533;
XX	
DT	05-SEP-2002 (first entry)
DE	
XX	Human corneal N-acetylgucosamine-6-sulfotransferase DNA SEQ ID NO:38.
KW	Human; N-acetylgucosamine-6-sulfotransferase; enzyme; GUCNA65T;
KW	cornea; Sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX	ophthalmologically; chromosome 16q22; gene; ds.
OS	Homo sapiens.
XX	
PM	US2002061562-A1.
XX	
ED	23-MAY-2002.
XX	
FE	09-AUG-2001; 2001US-0927602.

```
XX 11-AUG-2000; 2000US-32573P.
PR (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
XX FUKUDA MN, Akama TO;
XX
XX WPI: 2002-507643/54.
XX
XX New nucleic acid encoding corneal
PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
PT monitoring and diagnosis of macular corneal dystrophy -
XX
XX Claim 35: Page 31-53; 69pp; English.
XX
XX The present sequence encodes human corneal
CC N-acetylglucosamine-6-sulfotransferase (GlcNAc6ST) (I), which is able to
CC catalyze sulfation of keratan sulfate (KS). Also described is a method
CC for monitoring the effect of treatments for macular corneal dystrophy
CC (MCD), and detecting susceptibility to MCD. (I) is located to chromosome
CC 16q22, and has ophthalmological activity. (I) can be used to treat or
CC prevent macular corneal dystrophy types I or II. (I) makes possible
CC treatment of MCD without requiring keratoplasty or keratectomy.
XX
XX Sequence 48436 BP; 11653 A; 11904 C; 11645 G; 13234 T; 0 other:
SO
Query Match 2.9%; Score 59; DB 24; Length 48436;
Best Local Similarity 100.0%; Pred. No. 2,7e-15;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 678 AGGCTGGCGCTCCACAGCCAGCGTGCTCAGAGAGTGGCTTCTTCAACCTGCAG 736
DB 48024 AGGCTGGCGCTCCACAGCCAGCGTGCTCAGAGAGTGGCTTCTTCAACCTGCAG 48082
RESULT 18
AADO2697
ID AADO2697 standard; DNA; 160552 BP.
XX
AC AADO2697;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
XX
XX Human: glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KM therapy; selectin binding inhibitor; gene therapy; inflammation;
KM systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KM polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KM glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
KM Hashimoto's disease; Graves' disease; hypoparathyroidism; anaemia;
KM demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KM myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KM asthma; hypersensitivity; rheumatic fever; tissue rejection;
KM chromosome 16q23.1; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX 32847..32922
XX /tag= a
XX /number= 1
XX /label= 4a_5U4
XX 32923..35592
XX /tag= b
XX /cons-splice= (5'site:NO, 3'site:YES)
XX 35593..35674
XX /tag= c
XX /number= 2
XX /label= 4a_5U3
XX 35675..45093
XX /tag= d
XX intron
```

```
FT exon 45094..45185
FT /tag= e
FT /number= 3
FT /label= 4a_5U2
FT 45186..46633
FT /tag= f
FT /cons-splice= (5'site:NO, 3'site:NO)
FT 46634..46700
FT /tag= g
FT /number= 4
FT /label= 4a_5U1
FT 46701..47938
FT /tag= h
FT /cons-splice= (5'site:YES, 3'site:NO)
FT 47939..49746
FT /tag= i
FT /number= 5
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 47939..47935
FT /tag= j
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 47956..49128
FT /tag= k
FT /product= "Human glycosyl transferase-4alpha
FT (GST-4alpha)"
FT 49129..49746
FT /tag= l
FT 83257..83347
FT /tag= m
FT /label= 4a_5U2
FT 83348..96412
FT /tag= n
FT /cons-splice= (5'site:NO, 3'site:NO)
FT 96413..96484
FT /tag= o
FT /label= 4a_5U1
FT 96485..98456
FT /tag= p
FT /cons-splice= (5'site:NO, 3'site:NO)
FT 98457..99968
FT /tag= q
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 98457..98473
FT /tag= r
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 98474..99661
FT /tag= s
FT /product= "Human glycosyl transferase-4beta
FT (GST-4beta)"
FT 99662..99968
FT /tag= t
XX
XX WC200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000MO-US19741.
XX
XX 20-JUL-1999; 990S-0144694.
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX WPI: 2001-138471/14.
XX P-PSDB; AAY72639, AAY72640.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications -
XX
```

PS Example; Page 62-104; 128pp; English.

CC XX The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic  
CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on  
CC chromosome 16q23.1.  
CC GST is a type 2 membrane protein useful for inhibiting a binding event  
CC between a selectin and a selectin ligand, which comprises contacting the  
CC selectin with a non-sulphated selectin ligand, GST and a small molecular  
CC agent that inhibits the sulphation activity of GST. GST is also useful  
CC in inhibiting a selectin mediated binding event. GST is useful in gene  
CC therapy to treat disorders such as acute or chronic inflammation,  
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis  
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
CC glomerulonephritis, myocarditis, Sjogren's syndrome, Hashimoto's  
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious  
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,  
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
CC during transplantation.

XX

SQ Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

Query Match 2.9%; Score 59; DB 22; Length 160552;  
Best Local Similarity 100.0%; Pred. No. 2,5e-15;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0

DY 678 AGCGTCGCCCTCCTACGCACGAGTGGTCAAGAAGCTGCTTTCTTAACCTGAG 736  
Db 48446 AGCGTCGCCCTCCTACGCACGAGTGGTCAAGAAGCTGCTTTCTTAACCTGAG 48504  
|||||  
RESULT 19  
AAZ20798  
ID AAZ20798 standard; DNA; 37 BP.  
XX AAZ20798;  
XX  
XX  
DT 08-DEC-1999 (first entry)  
XX  
DE PCR primer for glycosyl sulfotransferase-3 coding sequence.  
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
KW Selectin binding interaction; inflammation; lymphocyte homing; human;  
KM secondary lymph organ; ss.  
XX  
XX Synthetic.  
OS  
FN WO9449018-A1.  
XX  
XX  
XX 30-SEP-1999.  
PD  
PF 26-FEB-1999; 99WO-US04316.  
XX  
XX 20-MAR-1998; 98US-0045284.  
PR 12-NOV-1998; 98US-0190511.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
PA (SYNT ) SYNTEX USA INC.  
XX  
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;  
DR WPI; 1999-580442/49.  
XX  
XX Human and murine glycosyl sulfotransferase 3 and related  
PT polynucleotides -  
XX  
PS Example 4; Page 30; 59pp; English.

CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of  
CC the invention. The nucleic acid sequences, probes and primers derived  
CC from these, proteins and antibodies are useful in detecting homologues.  
CC The sequences, antibodies and methods are useful in the diagnosis and

CC treatment of diseases associated with selectin binding interactions,  
CC including conditions associated with or resulting from the homing of  
CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
CC to secondary lymph organs.  
XX  
XX

SQ Sequence 37 BP; 14 A; 9 C; 8 G; 6 T; 0 other;

Query Match 1.8%; Score 37; DB 20; Length 37;  
Best Local Similarity 100.0%; Pred. No. 9.9e-06;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 884 AAACCTCAGAGGAGGACCACCCCTACTATGTGATGC 920  
Db 1 AAACCTCAGAGGAGGAGGACCACCCCTACTATGTGATGC 37  
|||||  
|||

RESULT 20  
ABK70821  
ID ABK70821 standard; DNA; 30 BP.  
XX  
XX ABK70821:  
AC  
XX 15-JUL-2002 (first entry)  
DT  
XX Probe for human gene CHST4.  
DE  
XX Human: sulphuric acid conjugation; ss; probe; CHST1; CHST3;  
XX KW CHST4; CHST5; CST; HNK-1ST; SULTA1; SULTB1; SULTX3; STE; TPST2.  
XX OS Homo sapiens.  
XX JP2002085067-A.  
EN  
XX 26-MAR-2002.  
PD  
XX 07-SEP-2000; 2000JP-0272229.  
PE  
XX 07-SEP-2000; 2000JP-0272229.  
PR  
XX (SAKA ) OTSUKA SEIYAKU KOGYO KK.  
PA  
XX WPI: 2002-378272/41.  
DR  
XX

Determination of enzymes participating in sulphuric acid conjugation in  
PT humans, useful for confirmation of safety of investigational drugs,  
PT comprises using oligonucleotide probes -  
XX  
XX Claim 4; Page 9; 13pp; Japanese.

The invention relates to classification and quantitative determination of  
CC enzymes participating in sulphuric acid conjugation comprising using  
CC oligonucleotide probes hybridising to the following regions: (a) 885-911  
CC region of CHST1 gene; (b) 174-197 region of CHST3 gene; (c) 1003-1032  
CC region of CHST4 gene; (d) 322-346 region of CHST5 gene; (e) 737-765  
CC region of CST gene; (f) 703-722 region of HNK-1ST gene; (g) 299-325  
CC region of SULTA1 gene; (h) 358-382 region of SULTB1 gene; (i) 534-582  
CC region of TPST2 gene; (j) 451-478 region of STE gene; and (k) 652-677  
CC region of SULTX3 gene. Also included are PCR primers for the above  
CC genes, kits and methods for determination. The probes, primers and the  
CC method are used in the determination of sulphuric acid conjugation for  
CC confirmation of the safety of investigational drugs. The present  
CC sequence is an oligonucleotide probe for one of the above listed genes.  
XX  
XX Sequence 30 BP; 5 A; 7 C; 7 G; 11 T; 0 other;

Query Match 1.5%; Score 30; DB 24; Length 30;  
Best Local Similarity 100.0%; Pred. No. 0.0096;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 1184 CGCTGCTCTTGCCCATGAAGAAGTTCT 1213  
|||||  
1 CGCTGCTCTTGCCCATGAAGAAGTTCT 30



RESULT 21  
AAZ20793  
ID AAZ20793 standard; DNA; 1926 BP.  
XX  
AC AAZ20793;  
XX  
DT 08-DEC-1999 (first entry)  
XX  
DE Mouse glycosyl sulfotransferase-3 coding sequence.  
XX  
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;  
KW secondary lymph organ; ss.  
XX  
OS Mus sp.  
XX  
PN M09949018-A1.  
XX  
PD 30-SEP-1999.  
XX  
PE 26-FEB-1999; 99WO-US04316.  
XX  
PR 20-MAR-1998; 98US-0045284.  
PR 12-NOV-1998; 98US-0190911.  
XX  
PA (REGC ) UNIV CALIFORNTA.  
PA (SYNT ) SYNTAX USA INC.  
XX  
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;  
XX  
DR WPI: 1999-580442/49.  
DR P-PSDB; AAY39919.  
XX  
PT Human and murine glycosyl sulfotransferase 3 and related  
PT polynucleotides -  
XX  
PS Claim 4; Fig 3; 59pp; English.  
XX  
CC This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of  
CC from these, proteins and antibodies are useful in detecting homologues.  
CC The sequences, antibodies and methods are useful in the diagnosis and  
CC treatment of diseases associated with selectin binding interactions,  
CC including conditions associated with or resulting from the homing of  
CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
CC to secondary lymph organs.  
XX  
SQ Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;  
XX  
Query Match 1.3%; Score 27; DB 20; Length 1926;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 494 GACATGACGCGCTTTGAGCCTACATG 520  
DB 714 GACATGACGCGCTTTGAGCCTACATG 740  
XX  
RESULT 22  
AAC74310  
ID AAC74310 standard; cDNA; 616 BP.  
XX  
AC AAC74310;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Human secreted protein gene 3i SEQ ID NO:41.  
XX  
KW Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antiinfective; antiproliferative; cytostatic; cardiac; vasotropic;  
KW cerebroprotective; neurotrophic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; vulnary; gene therapy; angiogenesis;  
DT

KW autoimmune disease; hyperproliferative disorder; infection; skin aging;  
KW wound healing; cardiovascular disorder; cerebrovascular disorder;  
KW nervous system disorder; food additive; preservative; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200057903-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 22-MAR-2000; 2000WO-US07525.  
XX  
PR 26-MAR-1999; 99US-0126595.  
PR 22-DEC-1999; 99US-0171549.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
DR WPI: 2000-584630/56.  
DR P-PSDB; AAB39340.  
XX  
PT New nucleic acid molecules encoding 48 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
PS Claim 1; Page 339; 395pp; English.  
XX  
CC The polynucleotide sequences given in AAC74280 to AAC74327 encode the  
CC human secreted proteins given in AAB39310 to AAB39357. AAB39358 to  
CC AAB39400 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;  
CC cytostatic; cardiac; vasotropic; cerebroprotective; neurotrophic;  
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
CC and vulnary. The polynucleotides and polypeptides can be used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They can also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 616 BP; 216 A; 102 C; 128 G; 169 T; 1 other;  
XX  
Query Match 1.1%; Score 23; DB 21; Length 616;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2010 CAATTAAGAAAAA 2032  
DB 570 CAATTAAGAAAAA 592  
XX  
RESULT 23  
AAS41008  
ID AAS41008 standard; cDNA; 1142 BP.  
XX  
AC AAS41008;  
XX  
DT 17-DEC-2001 (first entry)

PR	14-SEP-2000	2000US-023424.00
ER	14-SEP-2000	2000US-023363.00
PR	14-SEP-2000	2000US-023366.00
PR	14-SEP-2000	2000US-023365.54
PR	21-SEP-2000	2000US-023442.23
PR	21-SEP-2000	2000US-023442.74
PR	25-SEP-2000	2000US-023495.97
PR	25-SEP-2000	2000US-023498.00
PR	26-SEP-2000	2000US-023488.00
PR	27-SEP-2000	2000US-023583.64
PR	27-SEP-2000	2000US-023583.56
PR	29-SEP-2000	2000US-023632.27
PR	29-SEP-2000	2000US-023632.67
PR	29-SEP-2000	2000US-023635.68
PR	29-SEP-2000	2000US-023635.69
PR	02-OCT-2000	2000US-023658.02
PR	02-OCT-2000	2000US-023703.37
PR	02-OCT-2000	2000US-023703.38
PR	02-OCT-2000	2000US-023703.99
PR	02-OCT-2000	2000US-023740.00
PR	13-OCT-2000	2000US-023939.35
PR	13-OCT-2000	2000US-023939.37
PR	20-OCT-2000	2000US-024039.60
PR	20-OCT-2000	2000US-024112.21
PR	20-OCT-2000	2000US-024118.85
PR	20-OCT-2000	2000US-024178.85
PR	20-OCT-2000	2000US-024178.87
PR	20-OCT-2000	2000US-024180.08
PR	20-OCT-2000	2000US-024180.09
PR	20-OCT-2000	2000US-024182.66
PR	01-NOV-2000	2000US-024461.17
PR	08-NOV-2000	2000US-024645.74
PR	08-NOV-2000	2000US-024645.75
PR	08-NOV-2000	2000US-024654.76
PR	08-NOV-2000	2000US-024654.78
PR	08-NOV-2000	2000US-024655.23
PR	08-NOV-2000	2000US-024655.24
PR	08-NOV-2000	2000US-024655.25
PR	08-NOV-2000	2000US-024655.26
PR	08-NOV-2000	2000US-024655.27
PR	08-NOV-2000	2000US-024655.28
PR	08-NOV-2000	2000US-024655.32
PR	08-NOV-2000	2000US-024660.09
PR	08-NOV-2000	2000US-024661.11
PR	08-NOV-2000	2000US-024661.13
PR	17-NOV-2000	2000US-024929.07
PR	17-NOV-2000	2000US-024929.08
PR	17-NOV-2000	2000US-024929.09
PR	17-NOV-2000	2000US-024931.10
PR	17-NOV-2000	2000US-024931.12
PR	17-NOV-2000	2000US-024931.23
PR	17-NOV-2000	2000US-024931.24
PR	17-NOV-2000	2000US-024945.45
PR	17-NOV-2000	2000US-024945.46
PR	17-NOV-2000	2000US-024949.14
PR	17-NOV-2000	2000US-024921.15
PR	17-NOV-2000	2000US-024921.16
PR	17-NOV-2000	2000US-024921.17
PR	17-NOV-2000	2000US-024931.18
PR	17-NOV-2000	2000US-024931.19
PR	01-DEC-2000	2000US-025039.91
PR	05-DEC-2000	2000US-025103.00
PR	05-DEC-2000	2000US-025119.88
PR	05-DEC-2000	2000US-025617.19





KM		amplification; microorganism; cephalosporin; antibiotic; ds.
XX		
OS	Acremonium chrysogenum.	
XX		
FH	Location/Qualifiers	
FM	449..2198	/tag= a
FT	CDS	
FT		/product= O-acetylhomoserine sulphydrylase
FT	exon	449..495
FT		/tag= b
FT		/note= "exon 1"
FT	Intron	496..646
FT		/tag= c
FT		/note= "Intron 1"
FT	exon	647..701
FT		/tag= d
FT		/note= "exon 2"
FT	Intron	702..771
FT		/tag= e
FT		/note= "Intron 2"
FT	exon	772..1215
FT		/tag= f
FT		/note= "exon 3"
FT	Intron	1216..1315
FT		/tag= g
FT		/note= "Intron 3"
FT	exon	1316..1914
FT		/tag= h
FT		/note= "exon 4"
FT	Intron	1915..1980
FT		/tag= i
FT		/note= "Intron 4"
FT	exon	1981..2044
FT		/tag= j
FT		/note= "exon 5"
FT	Intron	2045..2105
FT		/tag= k
FT		/note= "Intron 5"
FT	exon	2106..2195
FT		/tag= l
FT		/note= "exon 6"
XX		
PN	JF08336391-A.	
PD		
XX	24-DEC-1996.	
XX		
PF	13-JUN-1995;	95JP-0145866.
XX		
PR	13-JUN-1995;	95JP-0145866.
PA	(ASAH ) ASAH KASEI KOGYO KK.	
XX		
DR	WPI: 1997-102718/10.	
DR	P-PSDB: AAW12400.	
PT	O-acetylhomoserine sulphydrylase gene from Acremonium chrysogenum -	
PT	useful for producing the antibiotic cephalosporin C	
XX		
PS	Claim 4; Page 12-14; 15pp; Japanese.	
XX		
CC	The sequence presented here is the genomic sequence of the Acremonium	
CC	chrysogenum O-acetylhomoserine sulphydrylase gene, more commonly known	
CC	as methionine synthase. The protein was isolated from A. chrysogenum	
CC	cells by conventional chromatographic methods and used for amino acid	
CC	sequencing. From peptide fragments of the protein, PCR primers	
CC	(AAT63304-5) were synthesised and used to amplify a fragment of the gene	
CC	for use as a probe to isolate the complete cDNA sequence (AAT63302).	
CC	The cDNA sequence was then used to isolate the genomic DNA sequence.	
CC	The gene can be used to transform microorganisms for the production of	
CC	cephalosporin C, a starting material for clinically important	
CC	cephalosporin type antibiotics.	
XX		
SQ	Sequence 2860 BP; 570 A; 933 C; 766 G; 591 T; 0 other;	

```

Query Match: 1.1%; Score 23; DB 18; Length 2860;
Best Local Similarity 100.0%; Pred.No. 7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 2010 CAAATAGAAAAA... 2032
|||||
DB 1248 CAAATAGAAAAA... 1226

RESULT 28
ABLI6362/c
ID ABLI6362 standard; DNA: 4871 BP.
XX
AC ABLI6362;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 559.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical; gene; ds.
XX
CS Drosophila melanogaster.
XX
FN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PI 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WP1: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 559; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI3051), expressed DNA
CC sequences (ABLI1840-ABLI6175) and the encoded proteins
CC (ABBI7737-ABBI2072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4871 BP; 1333 A; 1097 C; 1112 G; 1329 T; 0 other;

Query Match 1.1%; Score 23; DB 23; Length 4871;
Best Local Similarity 100.0%; Pred.No. 6.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 2010 CAAATAGAAAAA... 2032
|||||
DB 674 CAAATAGAAAAA... 652

RESULT 29
ABLI6364/c
ID ABLI6364 standard; DNA: 5101 BP.
XX
```

AC ABL16364;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 565.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
XX Claim 1: SEQ ID NO 565; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
XX sequences (ABB57737-ABB72072), and the encoded proteins  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 5101 BP; 1420 A; 1129 C; 1153 G; 1399 T; 0 other;  
SQ  
XX  
XX Query Match 1.1%; Score 23; DB 23; Length 5101;  
XX Best Local Similarity 100.0%; Pred. No. 6.8;  
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2010 CAAATTAAGAAAAA 2032  
DB 904 CAAATTAAGAAAAA 882  
RESULT 30  
ABL16360/C  
ID ABL16360 standard; DNA; 10138 BP.  
XX  
XX ABL16360;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 553.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD

XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
XX Claim 1: SEQ ID NO 553; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
XX sequences (ABB57737-ABB72072), and the encoded proteins  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 10138 BP; 3028 A; 2025 C; 2125 G; 2960 T; 0 other;  
SQ  
XX  
XX Query Match 1.1%; Score 23; DB 23; Length 10138;  
XX Best Local Similarity 100.0%; Pred. No. 6.5;  
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2010 CAAATTAAGAAAAA 2032  
DB 5941 CAAATTAAGAAAAA 5919  
RESULT 31  
ABK70834  
ID ABK70834 standard; DNA; 22 BP.  
XX  
XX ABK70834;  
XX  
XX 15-JUL-2002 (first entry)  
XX  
XX PCR primer for human gene CHST4 #1.  
DE  
XX  
XX Human; sulphuric acid conjugation; ss: PCR; CHST1; CHST3; primer;  
KW CHST4; CHST5; CST; HNK-1ST; SULT1A1; SULT1B1; SULT1X3; STE; TPST2.  
XX  
XX Homo sapiens.  
XX  
XX JP2002085067-A.  
XX  
XX 26-MAR-2002.  
XX  
XX 07-SEP-2000; 2000JP-0272229.  
XX  
XX 07-SEP-2000; 2000JP-0272229.  
XX  
XX (SAKA ) OTSUKA SEIYAKU KOGYO KK.  
XX  
XX WPI; 2002-378272/41.  
XX  
XX Determination of enzymes participating in sulphuric acid conjugation in  
XX humans, useful for confirmation of safety of investigational drugs,  
XX comprises using oligonucleotide probes -  
XX  
XX Claim 8: Page 10; 13pp; Japanese.  
PS

```

XX CC The invention relates to classification and quantitative determination of
CC enzymes participating in sulphuric acid conjugation comprising using
CC oligonucleotide probes hybridising to the following regions: (a) 885-911
CC region of CHST1 gene; (b) 174-197 region of CHST3 gene; (c) 1003-1032
CC region of CHST4 gene; (d) 322-346 region of CHST5 gene; (e) 737-765
CC region of CST gene; (f) 703-732 region of HNK-1ST gene; (g) 299-325
CC region of SULTR2L1 gene; (h) 358-382 region of SULTR1L1 gene; (i) 554-582
CC region of SULTR3 gene; (j) 451-478 region of STB gene; and (k) 652-677
CC region of TPST2 gene. Also included are PCR primers for the above
CC genes, kits and methods for determination. The probes, primers and the
CC method are used in the determination of sulphuric acid conjugation for
CC confirmation of the safety of investigational drugs. The present
CC sequence is a PCR primer for one of the above listed genes.
XX SQ Sequence 22 BP; 3 A; 8 C; 4 G; 7 T; 0 other:

Query Match          1.1%; Score 22; DB 24; Length 22:
Best Local Similarity 100.0%; Pred. No. 25;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1161 CCCTATATGCTCCAGGCTTG 1182
Db      1 CCCTATATGCTCCAGGCTTG 22

RESULT 32
AAS16960/C
ID AAS16960 standard; DNA; 22 BP.
XX AC AAS16960;
XX DT 12-MAR-2002 (first entry)
XX DE Human L-selectin sulfotransferase-2 (LST-2) cDNA RT-PCR primer #2.
XX XX Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
XX KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
XX KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
XX KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
XX KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
XX KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
XX KW antiallergic; PCR primer; reverse transcriptase.
XX OS Homo sapiens.
XX PN WO200185177-A1.
XX PD 15-NOV-2001.
XX PF 10-MAY-2001; 2001WO-US15452.
XX PR 11-MAY-2000; 2000US-0569320.
XX PA (BURN-) BURNHAM INST.
XX PI Fukuda M, Yeh J, Hiraoaka N;
XX DR WPI; 2002-075226/10.
XX PT New enzyme, useful for modifying acceptor molecule, comprises an
XX PT isolated L-selectin sulfotransferase-2 that directs expression of
XX PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
XX PT intestinal GlcNAc 6-sulfotransferase
XX PS Example 2; Page 49; 98pp; English.
XX CC The present invention provides a method of modifying an acceptor molecule
XX CC by contacting the acceptor with an isolated
XX CC beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
XX CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
XX CC invention also provides a method of treating or preventing an
XX CC L-selectin-mediated condition by reducing the expression or activity of a

```

```

CC beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3GNT,
CC and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3GNT. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to
CC clone cDNA encoding human LST-2.
XX SQ Sequence 22 BP; 6 A; 5 C; 7 G; 4 T; 0 other:

Query Match          1.1%; Score 22; DB 24; Length 22:
Best Local Similarity 100.0%; Pred. No. 25;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 ACAGTCACGCTTTCACTGG 559
Db      22 ACAGTCACGCTTTCACTGG 1

RESULT 33
AAS16962/C
ID AAS16962 standard; DNA; 22 BP.
XX AC AAS16962;
XX DT 12-MAR-2002 (first entry)
XX DE Human L-selectin sulfotransferase-2 (LST-2) cDNA RT-PCR primer #4.
XX XX Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
XX KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
XX KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
XX KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
XX KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
XX KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
XX KW antiallergic; PCR primer; reverse transcriptase.
XX OS Homo sapiens.
XX PN WO200185177-A1.
XX PD 15-NOV-2001.
XX PF 10-MAY-2001; 2001WO-US15452.
XX PR 11-MAY-2000; 2000US-0569320.
XX PA (BURN-) BURNHAM INST.
XX PI Fukuda M, Yeh J, Hiraoaka N;
XX DR WPI; 2002-075226/10.
XX PT New enzyme, useful for modifying acceptor molecule, comprises an
XX PT isolated L-selectin sulfotransferase-2 that directs expression of
XX PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
XX PT intestinal GlcNAc 6-sulfotransferase
XX PS Example 2; Page 49; 98pp; English.
XX CC The present invention provides a method of modifying an acceptor molecule
XX CC by contacting the acceptor with an isolated
XX CC beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
XX CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
XX CC invention also provides a method of treating or preventing an

```

CC L-selectin-mediated condition by reducing the expression or activity of a  
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done  
 CC by administering to the subject an oligosaccharide L-selectin antagonist  
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by  
 CC administering antibody material that specifically binds beta1,3gnt,  
 CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin  
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.  
 CC Alternatively, the expression or activity of LST-2 or its active  
 CC fragment can be reduced in combination with reducing the expression or  
 CC activity of beta1,3gnt. The method is useful for treating L-selectin  
 CC mediated conditions such as Crohn's disease and ulcerative colitis.  
 CC inflammatory disorders of the skin such as allergic contact dermatitis,  
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
 CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to  
 CC clone cDNA encoding human LST-2.

XX Sequence 22 BP: 5 A; 6 C; 5 G; 6 T; 0 other;

Query Match 1.1%; Score 22; DB 24; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 914 GTGATGAGCTCATCTGCCAA 935  
 DB 22 GTGATGAGCTCATCTGCCAA 1  
 |||

RESULT 34  
 AAS16964/c  
 ID AAS16964 standard; DNA: 22 BP.  
 XX AAS16964:  
 AC  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Human L-selectin sulfotransferase-2 (LST-2) cDNA RT-PCR primer #6.  
 XX  
 KW Human: beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;  
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;  
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;  
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;  
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antitumor;  
 KW antiinflammatory; antiproliferative; antidiabetic; dermatological;  
 KW antiallergic; PCR primer; reverse transcriptase.  
 XX  
 OS Homo sapiens.  
 PN WO200185177-A1.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 10-MAY-2001; 2001WO-US15452.  
 XX  
 PR 11-MAY-2000; 2000US-0569320.  
 XX  
 PA (BURN-) BURNHAM INST.  
 XX  
 PI Fukuda M, Yeh J, Hiraoaka N;  
 XX  
 DR WPI; 2002-075226/10.  
 XX  
 PT New enzyme, useful for modifying acceptor molecule, comprises an  
 PT isolated L-selectin sulfotransferase-2 that directs expression of  
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or  
 PT intestinal GlcNAc 6-sulfotransferase -  
 XX  
 PS Example 2; Page 49; 98pp; English.  
 XX  
 CC The present invention provides a method of modifying an acceptor molecule  
 CC by contacting the acceptor with an isolated  
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active  
 CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The

CC Invention also provides a method of treating or preventing an  
 CC L-selectin-mediated condition by reducing the expression or activity of a  
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done  
 CC by administering to the subject an oligosaccharide L-selectin antagonist  
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by  
 CC administering antibody material that specifically binds beta1,3gnt,  
 CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin  
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.  
 CC Alternatively, the expression or activity of LST-2 or its active  
 CC fragment can be reduced in combination with reducing the expression or  
 CC activity of beta1,3gnt. The method is useful for treating L-selectin  
 CC mediated conditions such as Crohn's disease and ulcerative colitis,  
 CC inflammatory disorders of the skin such as allergic contact dermatitis,  
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
 CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to  
 CC clone cDNA encoding human LST-2.

XX Sequence 22 BP: 8 A; 3 C; 8 G; 3 T; 0 other;

Query Match 1.1%; Score 22; DB 24; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1401 CTTCTGAGCTTGCTTGCCTACATCT 1422  
 DB 22 CTTCTGAGCTTGCTTGCCTACATCT 1  
 |||

RESULT 35  
 ABV38642/c  
 ID ABV38642 standard; cDNA: 341 BP.  
 XX ABV38642:  
 AC  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 38633.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene: ss.  
 XX  
 OS Homo sapiens.  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-198862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 7862; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;



CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 341 BP; 81 A; 76 C; 83 G; 101 T; 0 other;  
Query Match 1.1%; Score 22; DB 23; Length 341;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2011 AATAGAGAAAAA 2032  
D5 AATAGAGAAAAA 64  
DB 85 AATAGAGAAAAA 64  
RESULT 36  
ABV54232 standard; cDNA: 502 BP.  
ID ABV54232 standard; cDNA: 502 BP.  
XX  
XX ABV54232;  
XX  
XX 17-SEP-2002 (first entry)  
DE Human prostate expression marker cDNA 54223.  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200160860-A2.  
PN  
XX  
XX 23-AUG-2001.  
PD  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
PE  
XX  
XX 17-FEB-2000; 2000US-183319P.  
PR  
XX 16-MAR-2000; 2000US-189862P.  
PR  
XX 25-MAY-2000; 2000US-207454P.  
PR  
XX 09-JUN-2000; 2000US-211314P.  
PR  
XX 18-JUL-2000; 2000US-219007P.  
PR  
XX 13-DEC-2000; 2000US-255281P.  
PR  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
PI  
XX  
XX WPI: 2001-662795/76.  
DR  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
PT  
XX  
XX Claim 1; Page 10489; 11750pp; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 502 BP; 227 A; 87 C; 98 G; 90 T; 0 other;  
Query Match 1.1%; Score 22; DB 23; Length 502;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2011 AATAGAGAAAAA 2032  
D137 AATAGAGAAAAA 158  
DB 137 AATAGAGAAAAA 158  
RESULT 37  
AAS41556 standard; cDNA: 1781 BP.  
ID AAS41556 standard; cDNA: 1781 BP.  
XX  
XX AAS41556;  
XX  
XX 17-DEC-2001 (first entry)  
DE  
XX  
XX cDNA encoding novel human enzyme polypeptide #772.  
DE  
XX  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;  
KW anti arthritic; nephrotropic; anticoagulant; ss.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200155301-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX  
XX 17-JAN-2001; 2001WO-US01239.  
PE  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR  
XX 04-FEB-2000; 2000US-0180628.  
PR  
XX 24-FEB-2000; 2000US-0184664.  
PR  
XX 02-MAR-2000; 2000US-0186350.  
PR  
XX 16-MAR-2000; 2000US-0189874.  
PR  
XX 17-MAR-2000; 2000US-0190076.  
PR  
XX 18-APR-2000; 2000US-0198123.  
PR  
XX 19-MAY-2000; 2000US-0205515.  
PR  
XX 28-JUN-2000; 2000US-0209467.  
PR  
XX 28-JUN-2000; 2000US-0214886.  
PR  
XX 30-JUN-2000; 2000US-0215135.  
PR  
XX 07-JUL-2000; 2000US-0216647.  
PR  
XX 07-JUL-2000; 2000US-0216880.  
PR  
XX 11-JUL-2000; 2000US-0217487.  
PR  
XX 11-JUL-2000; 2000US-0217496.  
PR  
XX 14-JUL-2000; 2000US-0220890.  
PR  
XX 26-JUL-2000; 2000US-0220893.  
PR  
XX 26-JUL-2000; 2000US-0220964.  
PR  
XX 26-JUL-2000; 2000US-0220966.  
PR  
XX 14-AUG-2000; 2000US-0224518.  
PR  
XX 14-AUG-2000; 2000US-0224519.  
PR  
XX 14-AUG-2000; 2000US-0225213.  
PR  
XX 14-AUG-2000; 2000US-0225214.  
PR  
XX 14-AUG-2000; 2000US-0225266.  
PR  
XX 14-AUG-2000; 2000US-0225267.  
PR  
XX 14-AUG-2000; 2000US-0225268.  
PR  
XX 14-AUG-2000; 2000US-0225270.  
PR  
XX 14-AUG-2000; 2000US-0225447.  
PR  
XX 14-AUG-2000; 2000US-0225447.  
PR  
XX 14-AUG-2000; 2000US-0225757.  
PR  
XX 14-AUG-2000; 2000US-0225758.  
PR  
XX 14-AUG-2000; 2000US-0225759.  
PR  
XX 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241321.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0244674.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249266.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251859.  
 PR 08-DEC-2000; 2000US-0251988.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI, 2001:465566/50.  
 XX P-PSDB; AAU23686.  
 DR  
 DR  
 XX  
 XX  
 PT Novel polypeptides and polynucleotides useful for diagnosing,  
 PT preventing, treating neural, immune system, muscular, reproductive,  
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
 PT diseases  
 PI  
 XX  
 XX  
 Claim 4; SEQ ID No 782; 1180pp; English.  
 PS  
 PS  
 XX  
 CC The present invention relates to the isolation of novel human enzyme  
 CC polypeptides (AAU2915-AAU2814), and the cDNA and genomic sequences  
 CC encoding them. The enzyme polypeptides of the invention may comprise the  
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
 CC isomerases or ligases. The sequences of the invention are useful in the  
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
 CC disorders including hyperproliferative disorders (e.g. cancer),  
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
 CC blood-related disorders (e.g. haemophilia), reproductive disorders  
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
 CC polynucleotides of the invention can also be used in gene therapy.  
 CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human  
 CC enzyme polypeptides of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 CC  
 XX  
 XX  
 Query Match 1.1%; Score 22; DB 22; Length 1781;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2011 AATAGAGAAAAA 2032  
 |||  
 DB 1758 AATAGAGAAAAA 1779  
 |||  
 RESULT 38  
 ABR63555  
 ID ABR63555 standard; cDNA; 1782 BP.  
 XX  
 AC ABR63555;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Rat sequence differentially expressed in response to a hepatotoxin #1462.  
 KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
 KW differential expression; centrilobular necrosis; steatosis.  
 XX  
 OS Rattus norvegicus.  
 XX  
 PN WO200210453-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 30-JUL-2001; 2001WO-US23872.  
 XX  
 PR 31-JUL-2000; 2000US-222040P.  
 PR 02-NOV-2000; 2000US-244880P.  
 PR 11-MAY-2001; 2001US-290029P.  
 PR 15-MAY-2001; 2001US-290645P.  
 PR 22-MAY-2001; 2001US-292336P.  
 PR 06-JUN-2001; 2001US-295798P.  
 PR 13-JUN-2001; 2001US-297457P.  
 PR 19-JUN-2001; 2001US-298884P.  
 PR 09-JUL-2001; 2001US-303459P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff NR;  
 XX  
 DR WPI: 2002-241625/29.  
 XX  
 XX Predicting toxic effects of compounds or the progression of these toxic  
 PT effects by determining the changes in gene expression in tissues or  
 PT cells exposed to the toxin and comparing these to gene expression in  
 PT unexposed tissues or cells -  
 XX  
 PS Claim 1; Seq ID No 1462; 239pp; English.  
 XX  
 CC The invention relates to methods for predicting toxic effects of  
 CC compounds or the progression of these toxic effects by determining the  
 CC global changes in gene expression in tissues or cells exposed to the  
 CC toxin and comparing these to gene expression in unexposed tissues or  
 CC cells. Also included are methods of predicting at least one toxic  
 CC effect of a compound or progression of a toxic effect, preferably the  
 CC hepatotoxicity of a compound, comprising detecting the level of  
 CC expression in a tissue or cell sample exposed to the compound of two or  
 CC more genes listed in the specification, where differential expression of  
 CC the genes is indicative of at least one toxic effect or progression.  
 CC The method can also be used to identify an agent which modulates the  
 CC toxic response and predict cellular pathways that a compound modulates  
 CC in a cell. The methods utilise a set of at least two probes (on a solid  
 CC support in kit form), where each of the probes comprises a sequence that  
 CC specifically hybridises to a gene listed in the specification, a computer  
 CC system comprising a database containing information identifying the  
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
 CC set of genes comprising at least two genes listed in the specification,  
 CC and a user interface to view the information used to present information  
 CC identifying the expression level in a tissue or cell of at least one gene  
 CC listed in the specification. The method is useful for elucidating global  
 CC changes in gene expression and for identifying toxicity markers in  
 CC tissues or cell exposed to a known toxin. The genes may be used as  
 CC toxicity markers in drug screening and toxicity assays. The genes and

CC gene expression information may be used as diagnostic markers for the  
 CC prediction or identification of the physiological state of tissue or cell  
 CC sample that has been exposed to a compound or agent. Hepatotoxicity  
 CC is characterised by centrilobular necrosis and steatosis. The present  
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
 CC which is differentially expressed in response to a hepatotoxic agent.  
 XX  
 SQ Sequence 1782 BP; 413 A; 516 C; 474 G; 379 T; 0 other;  
 Query Match 1.1%; Score 22; DB 24; Length 1782;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2011 AATAGAGAAAAA 2032  
 |||  
 DB 1756 AATAGAGAAAAA 1777  
 |||  
 RESULT 39  
 AAH75633/C  
 ID AAH75633 standard; cDNA; 2415 BP.  
 XX  
 AC AAH75633;  
 XX  
 DT 31-OCT-2001 (first entry)  
 XX  
 DE Human ribosomal protein S4-19 encoding cDNA.  
 XX  
 KW Human; ribosomal protein S4-19; cytosolic; virucidal; immunomodulatory;  
 KW antiinflammatory; haemostatic; anti-HIV; malignant tumour; HIV;  
 KW human immunodeficiency virus; infection; immunological disease;  
 KW gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 1817..2356  
 FT CDS /tag= a  
 FT /product= "ribosomal protein S4-19"  
 FT /note= "Claimed in claim 6"  
 XX  
 PN WO200166581-A1.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PE 26-FEB-2001; 2001WO-CN00162.  
 XX  
 PR 07-MAR-2000; 2000CN-0111943.  
 XX  
 PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.  
 XX  
 PI Mao Y, Xie Y;  
 XX  
 DR WPI: 2001-565568/63.  
 DR P-PSDB; AAG77991.  
 PT  
 PT New human ribosomal protein S4-19 for diagnosing and treating malignant  
 PT tumour, haemopathy, human immunodeficiency virus infection, immunological  
 PT diseases and inflammations -  
 XX  
 PS Claim 6; Page 29-30; 35pp; Chinese.  
 XX  
 CC The invention relates to the human ribosomal protein S4-19 with  
 CC cytosolic, virucidal, immunomodulatory, antiinflammatory, haemostatic  
 CC and anti-HIV. The polypeptide and encoded polynucleotide are applicable  
 CC in diagnosis and treatment of malignant tumour, haemopathy, human  
 CC immunodeficiency virus (HIV) infection, immunological diseases and  
 CC various inflammation. The polynucleotide is useful for gene therapy. The  
 CC present sequence is that of the human ribosomal protein S4-19 encoding  
 CC cDNA.  
 XX  
 SQ Sequence 2415 BP; 611 A; 530 C; 599 G; 675 T; 0 other;

Query Match 1.1%; Score 22; DB 22; Length 2415;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2011 AATAGAGAAAAA 2032  
Db 1199 AATAGAGAAAAA 1178

## RESULT 40

AAQ11712/c  
ID AAQ11712 standard; DNA: 3138 BP.

XX AC AAQ11712;

XX DT 30-JUL-1991 (first entry)

XX DE Shuttle vector pMCM1630.

XX KM slime mould; replication; Rep gene; Ddp2; ss.

XX OS Dictyostelium discoideum.

XX FH Key Location/Qualifiers

XX FT sig\_peptide 2068 /\*tag= a  
XX FT /note= "finish position not given"

XX PN W09106644-A.

XX PD 16-MAY-1991.

XX PE 02-NOV-1990; 90WC-AU00530.

XX PR 02-NOV-1989; 89AU-0007187.

XX PA (UYMA-) MACQUARIE UNIV.

XX PI Slade MB, Chang ACM, Williams KL;

XX DR WPI: 1991-164194/22.

XX PT Polypeptide facilitating extra-chromosomal replication - of  
XX PS recombinant plasmid in Dictyostelium species

XX PS Claim 35; Fig 15; 90pp; English.

XX CC Shuttle vector pMCM1630 was constructed for use in E.coli and  
XX CC transfer back into Dictyostelium. It includes a fragment of the  
XX CC Dictyostelium plasmid Ddp2 containing the origin of replication,  
XX CC the promoter from the Dictyostelium Actin 15 gene, a sequence  
XX CC encoding the secretion signal peptide of the D19 gene of protein  
XX CC Psa, Actin 15 polyA signal and an ampicillin resistance-conferring  
XX CC sequence. See also AAQ11710 and AAQ11711.

XX SO Sequence 3138 BP; 979 A; 568 C; 574 G; 1017 T; 0 other;

Query Match 1.1%; Score 22; DB 12; Length 3138;  
Best Local Similarity 100.0%; Pred. No. 19;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2011 AATAGAGAAAAA 2032  
Db 2030 AATAGAGAAAAA 2009

## RESULT 41

ABA82684  
ID ABA82684 standard; DNA: 3597 BP.

XX AC ABA82684;

XX DT 25-JAN-2002 (first entry)

XX DE HSM800936 gene SEQ ID NO:70.

XX KM Human: high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;

XX KM sentence tagged site; STRs; osteoporosis; osteopathic; gene therapy;

XX KM antineuse therapy; vaccine; bone disorder; Paget's disease;

XX KM sclerostosis; osteomalacia; fibrous dysplasia; ds.

XX OS Homo sapiens.

XX PN W0200177327-A1.

XX PD 18-OCT-2001.

XX PF 21-JUN-2000; 2000WO-US16951.

XX PR 05-APR-2000; 2000US-0543771.

XX PR 05-APR-2000; 2000US-0544398.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Carulli JP, Little RD, Recker RR, Johnson ML;

XX DR WPI: 2001-657171/75.

XX PT New high bone mass (HBM) and Zmax1 genes and proteins useful for  
XX PS modulating bone mass for the treatment of e.g. osteoporosis -  
XX PS Claim 79; Page 372-373; 443pp; English.

XX CC The present invention describes the human Zmax1 gene and the high bone  
XX CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and  
XX CC HBM genes have osteopathic activities. The genes can be used in gene  
XX CC therapy, antineuse therapy and in the production of vaccines. They  
XX CC can be used in the diagnosis and treatment of bone disorders including  
XX CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous  
XX CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent  
XX CC sequences used in the exemplification of the present invention.

XX SO Sequence 3597 BP; 981 A; 929 C; 776 G; 911 T; 0 other;

Query Match 1.1%; Score 22; DB 22; Length 3597;  
Best Local Similarity 100.0%; Pred. No. 18;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2011 AATAGAGAAAAA 2032  
Db 3571 AATAGAGAAAAA 3592

## RESULT 42

ABLO7230/c  
ID ABLO7230 standard; CDNA: 73947 BP.

XX AC ABLO7230;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16172.

XX KM Drosophila: developmental biology; cell signalling; insecticide;

XX KM pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li FMD, Myers EM;  
 PI WPI: 2001-656860/75.  
 DR P-PSDB: ABB65127.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 16172; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA  
 CC sequences (AB16173-AB162072).  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 SO Sequence 73947 BP; 21410 A; 14389 C; 15751 G; 22397 T; 0 other:  
 Query Match 1.1%; Score 22; DB 23; Length 73947;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 22; Conservative 0; Indels 0; Gaps 0;  
 QY 2010 CAATTAAGAAAAA 2031  
 Db 73335 CAATTAAGAAAAA 73314  
 RESULT 43  
 ABK70835/c  
 ID ABK70835 standard; DNA: 21 BP.  
 XX  
 AC ABEK70835;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE PCR primer for human gene CHST4 #2.  
 XX  
 KM Human; sulphuric acid conjugation; ss; PCR; CHST1; CHST3; primer;  
 KM CHST4; CHST3; CST; HNK-1ST; SULT1A1; SULT1X3; STE; TPST2.  
 XX  
 OS Homo sapiens.  
 OS  
 PN JP2002085067-A.  
 PN  
 PD 26-MAR-2002.  
 PD  
 PF 07-SEP-2000; 2000JP-0272229.  
 PF  
 PR 07-SEP-2000; 2000JP-0272229.  
 PR  
 XX (SAKA ) OTSUKA SEIYAKU KOSYO KK.  
 PA  
 XX WPI: 2002-378272/41.  
 DR  
 XX  
 PT Determination of enzymes participating in sulphuric acid conjugation in  
 PT humans, useful for confirmation of safety of investigational drugs,  
 PT comprises using oligonucleotide probes -  
 XX  
 PS Claim 6; Page 11; 13pp; Japanese.  
 XX  
 CC The invention relates to classification and quantitative determination of  
 CC enzymes participating in sulphuric acid conjugation comprising using  
 CC oligonucleotide probes hybridising to the following regions: (a) 885-911

CC region of CHST1 gene; (b) 174-197 region of CHST3 gene; (c) 1003-1032  
 CC region of CHST4 gene; (d) 322-346 region of CHST5 gene; (e) 737-765  
 CC region of CST gene; (f) 703-732 region of HNK-1ST gene; (g) 299-325  
 CC region of SULT2A1 gene; (h) 358-382 region of SULT2B1 gene; (i) 554-582  
 CC region of SULT3 gene; (j) 451-478 region of STE gene; and (k) 652-677  
 CC region of TPST2 gene. Also included are PCR primers for the above  
 CC genes, kits and methods for determination. The probes, primers and the  
 CC method are used in the determination of sulphuric acid conjugation for  
 CC confirmation of the safety of investigational drugs. The present  
 CC sequence is a PCR primer for one of the above listed genes.  
 XX  
 SO Sequence 21 BP; 6 A; 6 C; 5 G; 4 T; 0 other:  
 Query Match 1.0%; Score 21; DB 24; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1240 CATGAATTTGCTGGCTACCG 1260  
 Db 21 CATGAATTTGCTGGCTACCG 1  
 RESULT 44  
 AAS16961  
 ID AAS16961 standard; DNA: 21 BP.  
 XX  
 AC AAS16961;  
 XX  
 DT 12-MAR-2002 (first entry)  
 DT  
 DE Human L-selectin sulfotransferase-2 (LST-2) cDNA RT-PCR primer #3.  
 DE  
 XX  
 KM Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;  
 KM L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;  
 KM ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;  
 KM allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;  
 KM delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;  
 KM antiinflammatory; antiproliferative; antidiabetic; dermatological;  
 KM antiallergic; PCR primer; reverse transcriptase.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200185177-A1.  
 PN  
 PD 15-NOV-2001.  
 PD  
 PF 10-MAY-2001; 2001WO-US15452.  
 PF  
 PR 11-MAY-2000; 2000US-0569320.  
 PR  
 XX (BURN-) BURHAM INST.  
 PA  
 XX Fukuda M, Yeh J, Hiraoaka N;  
 PI  
 XX WPI: 2002-075226/10.  
 DR  
 XX  
 PT New enzyme, useful for modifying acceptor molecule, comprises an  
 PT isolated L-selectin sulfotransferase-2 that directs expression of  
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or  
 PT intestinal glucNAc 6-sulfotransferase -  
 XX  
 PS Example 2; Page 49; 98pp; English.  
 PS  
 XX  
 CC The present invention provides a method of modifying an acceptor molecule  
 CC by contacting the acceptor with an isolated  
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active  
 CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The  
 CC invention also provides a method of treating or preventing an  
 CC L-selectin-mediated condition by reducing the expression or activity of a  
 CC beta1,3GNT that directs expression of a MECA-79 antigen. This can be done  
 CC by administering to the subject an oligosaccharide L-selectin antagonist  
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by  
 CC administering antibody material that specifically binds beta1,3GNT,

CC and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin  
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.  
CC Alternatively, the expression or activity of LSST-2 or its active  
CC fragment can be reduced in combination with reducing the expression or  
CC activity of beta1,3GNT. The method is useful for treating L-selectin  
CC mediated conditions such as Crohn's disease and ulcerative colitis,  
CC inflammatory disorders of the skin such as allergic contact dermatitis,  
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to  
CC clone cDNA encoding human LSST-2.

XX Sequence 21 BP; 0 A; 7 C; 6 G; 8 T; 0 other;

Query Match 1.0%; Score 21; DB 24; Length 21;  
Best Local Similarity 100.0%; Pred. No. 67;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 GGTCTGTCCTCTGGCGCTC 339  
DB 1 GGTCTGTCCTCTGGCGCTC 21

RESULT 45

AA27808/c  
ID AA27808 standard; DNA: 51 BP.

XX AA27808;

DT 24-JAN-2002 (first entry)

XX Human SNP oligonucleotide #1016.

XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
XX neuroprotective; antineurobiol; gene therapy; vaccine; amylose; cancer;  
XX amyloid protein; angiotensin; apoptosis related protein; cadherin;  
XX cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
XX complement related protein; cytochrome; kinase; cytokine; interferon;  
XX interleukin; G-protein coupled receptor; thioesterase; inflammation;  
XX multifactorial disease; autoimmune disease; infection;  
XX nervous system disease; ss.

XX Homo sapiens.

XX WC00147944-A2.

XX 05-JUL-2001.

XX 28-DEC-2000; 2000WC-US35498.

XX 28-DEC-1999; 9905-0173419.

XX 27-DEC-2000; 2000US-0173419.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI: 2001-465210/50.

PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
PT oncogenes and histones, useful for diagnosing and treating, e.g.

PT cancer, autoimmune diseases and infections -

XX Claim 1; Page 1670; 4143pp; English.

XX The present invention relates to oligonucleotides encoding polymorphic  
CC variants of proteins related to amylases, amyloid proteins, angiotensin,  
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
CC histones, kinases, colony stimulating factors, complement related  
CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,  
CC G-protein coupled receptors and thioesterases. The present sequence is  
CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
CC by them may be used in the prevention, diagnosis and treatment of

CC diseases associated with inappropriate expression of the proteins listed  
CC above. Disorders that may be prevented, diagnosed and/or treated include  
CC multifactorial diseases with a genetic component, such as autoimmune  
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
CC leukemia), diseases of the nervous system and an infection of pathogenic  
CC organisms.

XX Sequence 51 BP; 6 A; 4 C; 3 G; 38 T; 0 other;

Query Match 1.0%; Score 21; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 63;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATACAAAAA 2032  
DB 29 AATACAAAAA 9

RESULT 46

ABV08745/c  
ID ABV08745 standard; CDNA: 262 BP.

XX ABV08745;

DT 13-SEP-2002 (first entry)

XX Human prostate expression marker CDNA 8736.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200150860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-18319P.

XX 16-MAR-2000; 2000US-189862P.

XX 25-MAY-2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

XX 18-JUL-2000; 2000US-219007P.

XX 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI: 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 1378-1379; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;  
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 262 BP; 56 A; 56 C; 58 G; 88 T; 4 other;  
Query Match 1.0%; Score 21; DB 23; Length 262;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2012 AATAAGAAAAAAAAAAAAA 2032  
Db 33 AATAAGAAAAAAAAAAAAA 13  
RESULT 47  
AAF63356  
ID AAF63356 standard; DNA; 294 BP.  
XX  
AC AAF63356;  
XX  
DT 04-MAY-2001 (first entry)  
XX  
DE DNA encoding larval T. solium TSRS1 protein.  
XX  
KW Larval Taenia solium; cysticercosis; pork tapeworm; vaccine; TSRS1; ds.  
XX  
OS Taenia solium.  
XX  
FH Key Location/Qualifiers  
FT CDS 3..227  
FT /\*tag= a  
FT /product= "TSRS1"  
FT /partial  
XX  
FN WO200110697-A2.  
XX  
PD 15-FEB-2001.  
XX  
PF 03-AUG-2000; 2000WO-US21173.  
XX  
PR 05-AUG-1999; 99US-0147318.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Tsang VCM, Greene RM, Wilkins PP, Hancock K;  
XX  
DR WPI: 2001-202757/20.  
DR P-PSDB: AAB72214.  
XX  
PT Composition for detecting larval Taenia solium, comprising isolated,  
PT synthetic or recombinant larval Taenia solium polypeptides that are  
PT immunoreactive with Taenia solium antibodies  
XX  
PS Claim 7; Page 34-35; 37pp; English.  
XX  
CC This invention relates to a composition comprising one or more isolated,  
CC synthetic or recombinant larval Taenia solium polypeptides, or its  
CC antigenic fragments. The invention includes T. solium polypeptides TS-14,  
CC TS-18 and TSRS-1 and their encoding polynucleotide sequences, also  
CC included are T. solium antigenic peptides. A T. solium polypeptide is  
CC useful for diagnosing cysticercosis in a mammal. T. solium is also  
CC useful for detecting T. solium antibodies in a biological sample. The  
CC proteins are useful for reducing, possibly preventing, T. solium  
CC infection or transmission and are used in immunoassays for detecting  
CC T. solium. Nucleic acid sequences encoding the proteins are useful as  
CC molecular probes or primers for detecting RNA and DNA involved  
CC transcription and translation of the proteins. The proteins and  
CC polynucleotide sequences may be used to form a vaccine which will protect  
CC against T. solium also known as pork tapeworm. The present sequence  
CC represents DNA encoding TSRS1.  
XX  
SQ Sequence 294 BP; 107 A; 50 C; 71 G; 66 T; 0 other;

Query Match 1.0%; Score 21; DB 22; Length 294;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2012 AATAAGAAAAAAAAAAAAA 2032  
Db 271 AATAAGAAAAAAAAAAAAA 291  
RESULT 48  
ABV56921/C  
ID ABV56921 standard; cDNA; 397 BP.  
XX  
AC ABV56921;  
XX  
DT 17-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 56912.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI: 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
PS Claim 1; Page 10964; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213)/of the  
CC specification or its complement. (1) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 397 BP; 75 A; 89 C; 87 G; 146 T; 0 other;  
Query Match 1.0%; Score 21; DB 23; Length 397;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2012 AATAAGAAAAAAAAAAAAA 2032  
|||||

D5 44 AATTAGCAAAAAAAAAA 24

Search completed: January 10, 2003, 21:45:33  
Job time : 965 secs



Mon Jan 13 09:13:58 2003

us-09-816-825-1.ol120.rst

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 21:44:02 ; Search time 2885 Seconds

(without alignments)  
11407.012 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032

Sequence: 1 ggcctcagagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO\_NUC  
Gapox 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 20

Total number of hits satisfying chosen parameters: 1660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result. %  
No. Score Match Length DB ID Description  
SUMMARIES  
1 630 31.0 954 13 B1823850 603039012  
2 581 28.6 593 14 BM969292  
3 551 27.1 571 13 BM129080  
4 478 23.5 553 13 BM128370  
5 478 23.5 553 13 BM128831  
6 468 23.0 668 9 AL709927  
AL709927 DKFZp6860

7 318 15.6 680 12 BE878439  
8 288 14.2 362 10 AM002418  
9 283 13.9 419 10 AM572390  
10 150 7.4 417 10 AM572510  
11 136 6.7 358 12 BE056840  
12 59 2.9 525 12 BE197521  
13 59 2.9 620 9 A1824100  
14 59 2.9 695 12 BE857538  
15 59 2.9 783 13 BG663298  
16 47 2.3 436 10 AM081348  
17 27 1.3 497 13 BM246681  
18 27 1.3 536 13 BM245312  
19 27 1.3 695 13 BG664671  
20 27 1.3 783 13 BG663298  
21 27 1.3 852 13 BG666340  
22 27 1.3 1923 11 AK009113  
23 25 1.2 501 13 BM036473  
24 25 1.2 682 17 CNO02BKQ  
25 24 1.2 375 13 BM663231  
26 24 1.2 566 13 BM666130  
27 24 1.2 919 17 CNO01L7Y  
28 24 1.1 132 17 B75634  
29 23 1.1 191 9 A1072478  
30 23 1.1 269 10 B1828285  
31 23 1.1 304 14 BQ192511  
32 23 1.1 319 10 BQ192511  
33 23 1.1 342 12 BG291580  
34 23 1.1 349 13 B1555949  
35 23 1.1 362 10 BE340900  
36 23 1.1 385 14 BM930440  
37 23 1.1 397 13 B1505360  
38 23 1.1 408 10 BE109147  
39 23 1.1 413 9 A1095726  
40 23 1.1 419 10 AM080577  
41 23 1.1 420 10 BE107411  
42 23 1.1 423 9 A1195993  
43 23 1.1 437 14 W32300  
44 23 1.1 443 9 A1952475  
45 23 1.1 446 14 BQ013210  
46 23 1.1 461 14 BQ427521  
47 23 1.1 481 14 BQ194968  
48 23 1.1 484 10 AM434101  
49 23 1.1 484 12 BE062672  
50 23 1.1 497 9 AA809972  
51 23 1.1 498 13 BM316129  
52 23 1.1 501 9 AA765670  
53 23 1.1 538 10 BE419040  
54 23 1.1 545 13 B1407057  
55 23 1.1 577 12 BG897008  
56 23 1.1 616 17 A2424759  
57 23 1.1 616 10 BE109266  
58 23 1.1 626 17 AC183235  
59 23 1.1 720 14 BQ772892  
60 23 1.1 721 14 BQ775707  
61 23 1.1 819 12 BE875409  
62 23 1.1 834 14 BQ884877  
63 23 1.1 857 13 B1517608  
64 23 1.1 1048 13 BM93197  
65 23 1.1 1048 13 BM935024  
66 23 1.1 1068 13 CNO02BKQ  
67 22 1.1 66 17 CNO02BKQ  
68 22 1.1 67 14 BQ395065  
69 22 1.1 77 9 A1312401  
70 22 1.1 86 10 AA0409770  
71 22 1.1 93 10 AM055842  
72 22 1.1 93 10 AV742926  
73 22 1.1 96 13 B1697249  
74 22 1.1 135 13 B1404524  
75 22 1.1 139 17 A2896024  
76 22 1.1 152 13 B1268683  
77 22 1.1 154 13 B1042721  
78 22 1.1 155 10 BE655780  
79 22 1.1 155 13 BM187070  
156 9 AU037435  
AU037435

BE878439 MR0-ET010  
AM002418 wu61503.x  
AM572390 xu09a06.x  
AM572510 xq18g11.x  
BE056840 7x10d12.x  
BE197521 7084d08.x  
A1824100 w146c01.x  
BE857538 7901a08.x  
BE858652 7901a09.x  
AM081348 xc41b06.x  
BM246681 K0741B04-  
BM245312 K0722H07-  
BG664671 602831875  
BG663298 602827716  
BG666340 602832826  
AK009113 Mus. muscu  
BM036473 fu81f07.x  
A1189971 Tetraodon  
BM663231 UT-E-DX1-  
BM666130 UT-E-DX1-  
A1173359 Tetraodon  
B75634 PC111-1M1  
A1072478 UT-R-C2-n  
B1828285 B1828285  
BQ192511 UT-R-DX1-  
BE117198 UT-R-B51-  
BG291580 6028185671  
B1555949 603339076  
BE340900 EST344988  
BM930440 UT-E-B11-  
B1505360 B170026A  
BE109147 UT-R-B51-  
A1095726 qb24112.x  
AM080577 xc3se07.x  
BE107411 UT-R-B51-  
A1195993 u163b09.x  
W32300 zc05q05.81  
A1952475 wx74604.x  
BQ013210 UT-1-BC1P  
BQ427521 AGENCOURT  
BQ194968 UT-R-CM1-  
AM434101 UT-R-BJOP  
BE062672 7h63g02.x  
AA809972 ca95f12.s  
BM316129 fw65g11.x  
AA765670 ca04f06.s  
BE419040 WMR018.C5  
B1407057 189A02.Ma  
BG897008 HOA59-1-H  
A2424759 1M0204G04  
BE109266 UT-R-B51-  
AC183235 Pan. trogl  
BQ772892 UT-H-EB0-  
BQ775707 UT-H-FH0-  
BE875409 601489128  
BQ884877 AGENCOURT  
B1517608 603042215  
BM93197 50071-2-7  
BM935024 50072-2-7  
CNO02BKQ Tetraodon  
A1210258 Tetraodon  
BQ395065 NISC-ng12  
A1312401 ta80e08.x  
AA0409770 fd02c07.y  
AM055842 1d83f09.y  
AV742926 AV742926  
B1697249 603446757  
B1404524 MT-E-CP1-  
A2896024 RPCI-24-1  
B1268683 NF01810G  
B1042721 CM4-OT016  
BE655780 UT-W-AJ1-  
BM187070 fv81f11.y  
AU037435 AU037435











C 810	20	1.0	122	9	A1872554	WM75907.x	C 883	20	1.0	254	9	A1189443	q0d2809.x
C 811	20	1.0	122	10	BE682760		C 884	20	1.0	254	12	BF653319	277122.MA
C 812	20	1.0	131	10	BF387908	UI-R-CAL-	C 885	20	1.0	254	13	BM150024	TCAPAD11
C 813	20	1.0	132	14	BQ593635	PFEStoab1	C 886	20	1.0	255	9	AA374048	ESTB6320
C 814	20	1.0	136	9	AT036206	uh20c12.x	C 887	20	1.0	255	10	AV783156	AV783156
C 815	20	1.0	139	9	AT131969	ta89e01.x	C 888	20	1.0	260	13	BM095063	sa125h07.
C 816	20	1.0	141	9	A1250015	qrx24905.x	C 889	20	1.0	261	9	AA380688	EST3927.
C 817	20	1.0	144	14	BQ265053	NISC-L103	C 890	20	1.0	261	10	AM534716	UI-R-B50-
C 818	20	1.0	146	13	BM534103	tw93909.y	C 891	20	1.0	262	9	A1583642	tl74d04.x
C 819	20	1.0	148	10	BE071413	RC2-HT050	C 892	20	1.0	262	9	AA299688	EST12233
C 820	20	1.0	148	12	BF002054	7498d05.x	C 893	20	1.0	262	9	AA433767	h02b12.s
C 821	20	1.0	151	17	AZ257451	RPCI1-23-1	C 894	20	1.0	262	13	BM144357	TCAPAD110
C 822	20	1.0	157	9	AA198802	mm1d10.x	C 895	20	1.0	265	10	BE317804	NFO70C12L
C 823	20	1.0	158	13	BF144339	372786.MA	C 896	20	1.0	268	13	BM642814	BS3200500
C 824	20	1.0	159	12	BF418502	UI-R-BJ2-	C 897	20	1.0	268	13	BM644966	BS3200080
C 825	20	1.0	160	9	A1345714	tb84c09.x	C 898	20	1.0	271	14	BM774160	ra90b07.y
C 826	20	1.0	161	9	A1252073	qv58a10.x	C 899	20	1.0	272	14	L38218	BNAF1100E.M
C 827	20	1.0	162	12	BG654215	sa653c05.	C 900	20	1.0	273	10	AM487664	89994.MAR
C 828	20	1.0	164	9	A1836277	UI-M-A00-	C 901	20	1.0	277	12	BF581312	602100587
C 829	20	1.0	164	14	BQ739491	PFEStoab4	C 902	20	1.0	278	9	AA122813	tm03810.r
C 830	20	1.0	165	9	A1342993	qoc27e08.x	C 903	20	1.0	278	10	BM443348	BB445348
C 831	20	1.0	166	13	BM157733	tw26f05.y	C 904	20	1.0	279	10	AM524153	UI-R-B00-
C 832	20	1.0	168	10	AM292058	UI-H-B12-	C 905	20	1.0	280	13	BM089829	503682.MA
C 833	20	1.0	169	10	AV523525	AV523525	C 906	20	1.0	281	14	BQ769300	PFEStoab1
C 834	20	1.0	174	10	AM273389	xx38f09.x	C 907	20	1.0	282	12	BE991525	UI-R-B21-
C 835	20	1.0	174	10	AM551273	10076B08-	C 908	20	1.0	282	12	BF992125	UI-M-B21-
C 836	20	1.0	176	10	AM275671	xx39f05.x	C 909	20	1.0	283	14	C91311	C91311
C 837	20	1.0	177	9	A1434121	tl31d06.x	C 910	20	1.0	286	10	AM103667	x682c07.x
C 838	20	1.0	177	9	AL697773	DKFZP6860	C 911	20	1.0	286	12	BF704545	MI-P-H4-a
C 839	20	1.0	179	12	BG182541	RG12541.RST14.17.A	C 912	20	1.0	288	9	A1680307	tw62c02.x
C 840	20	1.0	182	9	AU057285	AU057285	C 913	20	1.0	289	12	BG382642	298614.MA
C 841	20	1.0	183	9	AA808666	cb38e09.s	C 914	20	1.0	289	14	BM029517	UI-R-EJ1-
C 842	20	1.0	186	9	AA433304	TEUPF0127	C 915	20	1.0	290	9	A1739307	w130c11.x
C 843	20	1.0	186	10	BE120123	UI-R-CA0-	C 916	20	1.0	290	12	BF412860	UI-R-BT1-
C 844	20	1.0	187	13	BQ080118	BQ080118	C 917	20	1.0	290	13	BM540016	hb17a08.9
C 845	20	1.0	188	12	BF781021	60210650	C 918	20	1.0	290	14	BQ184202	UI-E-EJ1-
C 846	20	1.0	190	12	BG627020	CC-ess1cL	C 919	20	1.0	293	10	AM150288	xx37c08.x
C 847	20	1.0	193	12	BG361584	pb55a08.y	C 920	20	1.0	294	12	BG372147	ra163d10.
C 848	20	1.0	193	13	BM144239	TCAPAD110	C 921	20	1.0	295	9	AA210268	mu71b10.r
C 849	20	1.0	197	12	BF512879	UI-H-B13-	C 922	20	1.0	296	9	A1915281	xx06f03.x
C 850	20	1.0	200	14	BQ596238	PFEStoab3	C 923	20	1.0	296	14	N96010	2033c3.c2AP
C 851	20	1.0	204	13	BI494079	d110f609.	C 924	20	1.0	297	14	BQ194687	UI-R-CN1-
C 852	20	1.0	202	13	BQ18521	BQ18521	C 925	20	1.0	298	17	BH426986	BOHME29TF
C 853	20	1.0	207	10	AV366356	AV366356	C 926	20	1.0	300	9	AL037346	MEB01A01
C 854	20	1.0	207	10	AM122482	UI-M-BH2-	C 927	20	1.0	300	9	AU057349	AU057349
C 855	20	1.0	212	9	A1454761	UI-R-C2P-	C 928	20	1.0	301	9	AA996895	UI-R-C0-h
C 856	20	1.0	212	13	BG941967	ax19e07.x	C 929	20	1.0	301	14	BM125882	k04f905.y
C 857	20	1.0	217	13	BI508797	BI170008A	C 930	20	1.0	303	10	BE108291	UI-R-B51-
C 858	20	1.0	217	13	BM193679	TCAPAD113	C 931	20	1.0	303	12	BG626902	CC-ess1cL
C 859	20	1.0	217	14	BQ577165	PFEStoabl	C 932	20	1.0	304	9	A1282038	q988e10.x
C 860	20	1.0	218	12	BG571976	602592430	C 933	20	1.0	304	13	BI508033	BI170006B
C 861	20	1.0	218	14	H21316	AS31SB231SL	C 934	20	1.0	306	12	BF403103	UI-R-CA0-
C 862	20	1.0	221	9	A1548904	UI-R-C3-t	C 935	20	1.0	307	9	AU057452	UI-R-C0-h
C 863	20	1.0	221	12	BE990056	UI-M-B21-	C 936	20	1.0	307	17	CN057452	ANOpheles
C 864	20	1.0	224	9	BB589593	BB589593	C 937	20	1.0	308	10	BE156860	RCL-HT037
C 865	20	1.0	228	9	A1342140	q126d06.x	C 938	20	1.0	308	14	BQ754571	EBec01-SO
C 866	20	1.0	228	14	BM812127	IX14f01.y	C 939	20	1.0	308	14	BQ770646	UI-M-F10-
C 867	20	1.0	231	12	BF389709	UI-R-B52-	C 940	20	1.0	310	9	AL373463	AL373463
C 868	20	1.0	233	13	BI849908	477882.MA	C 941	20	1.0	311	13	BI153086	GM21156.5
C 869	20	1.0	235	10	AM397585	sg81f04.y	C 942	20	1.0	311	14	BO152492	NFO70C051
C 870	20	1.0	241	10	BE156861	RCL-HT037	C 943	20	1.0	312	14	BO152492	NFO19C051
C 871	20	1.0	242	9	AV263129	AV263129	C 944	20	1.0	313	14	BQ393435	NISC-ng03
C 872	20	1.0	243	10	AM452160	UI-H-B13-	C 945	20	1.0	316	10	AM567190	AV567190
C 873	20	1.0	243	10	AM493776	UI-M-BH3-	C 946	20	1.0	316	10	AM291148	UI-H-B12-
C 874	20	1.0	244	10	BE057197	sm98a11.y	C 947	20	1.0	317	9	A1448797	ms39h08.x
C 875	20	1.0	246	9	A1031183	5MTP.Xemo	C 948	20	1.0	317	9	A1448797	ms39h08.x
C 876	20	1.0	248	10	BE201696	FK97D11.x	C 949	20	1.0	320	9	AA779839	a145e09.s
C 877	20	1.0	249	9	AT483140	v136c06.x	C 950	20	1.0	320	10	AV738188	AV738188
C 878	20	1.0	249	12	BE955637	UI-M-C0G0	C 951	20	1.0	321	10	AM168414	x182h01.x
C 879	20	1.0	249	17	N97395	0562M7.gmbp	C 952	20	1.0	321	12	BG290837	602389179
C 880	20	1.0	251	10	AM520963	UI-R-B00-	C 953	20	1.0	322	10	AM122680	UI-M-BH2-
C 881	20	1.0	251	14	BQ936880	AGENCOURT	C 954	20	1.0	323	9	A1333674	q908h04.x
C 882	20	1.0	253	9	A1577328	UI-R-Y0-v	C 955	20	1.0	324	9	AA643802	np10e06.s

C 956	20	1.0	324	13	BI398805
C 957	20	1.0	324	14	R61857
C 958	20	1.0	327	9	A1090983
959	20	1.0	328	10	AM104967
960	20	1.0	328	10	BE583045
C 961	20	1.0	328	13	BI286903
C 962	20	1.0	328	9	A1434280
C 963	20	1.0	329	10	BE206623
C 964	20	1.0	329	14	BO633533
C 965	20	1.0	331	9	AA456748
966	20	1.0	332	9	AA909064
C 967	20	1.0	332	10	AV801215
C 968	20	1.0	332	17	A2573716
C 969	20	1.0	332	9	A1241046
C 970	20	1.0	334	14	BQ451509
C 971	20	1.0	335	13	BM275955
C 972	20	1.0	335	17	AO903164
C 973	20	1.0	336	9	AA964643
C 974	20	1.0	336	9	A1285657
C 975	20	1.0	336	9	A1715385
C 976	20	1.0	336	13	BM276237
C 977	20	1.0	338	9	A1043582
C 978	20	1.0	339	9	A1157441
C 979	20	1.0	339	10	BE196907
C 980	20	1.0	339	13	BI290968
C 981	20	1.0	340	14	BO739631
C 982	20	1.0	342	9	AA865498
C 983	20	1.0	342	10	AW305061
C 984	20	1.0	344	9	AA737951
C 985	20	1.0	344	10	BE097550
C 986	20	1.0	345	13	BM193812
987	20	1.0	346	10	AA440511
C 988	20	1.0	347	12	BE864396
C 989	20	1.0	347	13	BM572191
C 990	20	1.0	348	9	AA309612
C 991	20	1.0	349	13	BI505182
C 992	20	1.0	349	14	BM718552
993	20	1.0	350	9	A1344948
C 994	20	1.0	350	12	BG017440
995	20	1.0	350	12	BF388844
C 996	20	1.0	352	9	A1028968
C 997	20	1.0	353	9	AL657651
998	20	1.0	354	13	BI708363
C 999	20	1.0	357	10	AW243036
C1000	20	1.0	357	9	AA827209

## ALIGNMENTS

RESULT 1  
 BI823850  
 LOCUS  
 DEFINITION  
 603039012F1 NIH\_MGC\_115 Homo sapiens CDNA clone IMAGE:1179826 5',  
 mRNA sequence.  
 BI823850  
 BI823850.1 GI:15935400  
 EST.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 954)  
 NIH-MGC, <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1448 row: h column: 03  
 High quality sequence stop: 856.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..954  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1179826"  
 /clone\_lib="NIH\_MGC\_115"  
 /lab\_host="DH10B"  
 /note="Organ: Pooled brain, lung, testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age 69. Library is male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

BASE COUNT 199 a 297 c 250 g 208 t  
 ORIGIN  
 Query Match 31.0%; Score 630; DB 13; Length 954;  
 Best Local Similarity 100.0%; Pred. No. 11e-186;  
 Matches 630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 CAAGTCTTCCACTGACCAATGCTACTGCTCTTAAATGAAGCTCTGCTTTCT 219  
 |||||  
 DB 56 CAGGCTTCCACTGACCAATGCTACTGCTCTTAAATGAAGCTCTGCTTTCT 115  
 QY 220 GGTTCCTCCAGATGGCCATGTCGCTATTTCCATGATGACGACCAATGATGCTC 279  
 |||||  
 DB 116 GGTTCCTCCAGATGGCCATGTCGCTATTTCCATGATGACGACCAATGATGCTC 175  
 QY 280 CTTGTATGAAAGGACAGCCGAGCGATGACGCTGCTGCTCTTCTCTGCGCTC 339  
 |||||  
 DB 176 CCGTGTATGAAAGGACAGCCGAGCGATGACGCTGCTGCTCTTCTCTGCGCTC 235  
 QY 340 TGGCTCTCTTTTGGGGCAGCTTTTGGGAGACCCAGATGTTTCTACTGATGA 399  
 |||||  
 DB 236 TGGCTCTCTTTTGGGGCAGCTTTTGGGAGACCCAGATGTTTCTACTGATGA 295  
 QY 400 GCCCGCTGGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 459  
 |||||  
 DB 296 GCCCGCTGGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 355  
 QY 460 TGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 519  
 |||||  
 DB 356 TGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 415  
 QY 520 GGAACCTGTCCTCCGAGACAGTCAGCTCTTCACTGAGGAGACAGCCGCTGTG 579  
 |||||  
 DB 416 GGAACCTGTCCTCCGAGACAGTCAGCTCTTCACTGAGGAGACAGCCGCTGTG 475  
 QY 580 TTCTGACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 639  
 |||||  
 DB 476 TTCTGACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 535  
 QY 640 GCTGCTGTCAGACCAACACCTTGGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 699  
 |||||  
 DB 536 GCTGCTGTCAGACCAACACCTTGGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 595  
 QY 700 CGTGTGTCTCAAG 759  
 |||||  
 DB 596 CGTGTGTCTCAAG 655  
 QY 760 CCCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATG 769  
 |||||  
 DB 656 CCCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATG 685



RESULT 2  
BM969292/c 593 bp mRNA linear EST 21-MAR-2002  
LOCUS  
DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone  
ACCESSION BM969292  
VERSION BM969292.1 GI:19586879  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 593)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL  
MEDLINE 9704447  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.  
FEATURES  
SOURCE  
Location/Qualifiers  
1..593  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-CF-ENO-acp-1-21-0-UI"  
/clone\_id="UI-CF-ENO"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Ecor I; Site\_2: Not I;  
UI-CF-ENO is a cDNA library containing the following  
tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells.  
The library was constructed according to Bonaldo, Lennon  
and Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an Ecor I adaptor, digested with Not I, and cloned  
directionally into pT73-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (drr)18 tail. The sequence tag for this  
library is CTGCTCAGGT.  
TAG\_LIB-UI-CF-ENO  
TAG-TISSUE-Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG\_SEQ-CTGCTCAGGT"

BASE COUNT 157 a 127 c 137 g 172 t  
ORIGIN

Query Match 28.6%; Score 581; DB 14; Length 593;  
Best Local Similarity 100.0%; Pred. No. 2.7e-171;  
Matches 581; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 1436 ACATCTGTGGGTATCAGACTGAGTGTGTCACAGCTCAGCAAGGA 1495  
|||||  
DB 593 ACATCTGTGGGTATCAGACTGAGTGTGTCACAGCTCAGCAAGGA 534  
OY 1436 CTTTGTGTCATGCTTGTGTAGAAAAACAGACTGGGAACTTATGTGAGCAGACAT 1555

|||||  
Db 533 CTTTGTGTCATGCTTGTGTAGAAAAACAGACTGGGAACTTATGTGAGCAGACAT 474  
OY 1556 CCCACAGTGAACAGAGATATGCTCTTCTTTCTTGATCTTCTGAGCAGAC 1615  
|||||  
Db 473 CCCACAGTGAACAGAGATATGCTCTTCTTTCTTGATCTTCTGAGCAGAC 414  
OY 1616 TTGAGACTTTGTGGCTTGGAGGCTATTAGCAGACAGATATCAGTGGATGATC 1675  
|||||  
Db 413 TTGAGACTTTGTGGCTTGGAGGCTATTAGCAGACAGATATCAGTGGATGATC 354  
OY 1676 CATAACTCCCTGTCACATCTTGCCCAATGGGATGATTTCCACCAAGAGCTCA 1735  
|||||  
Db 353 CATAACTCCCTGTCACATCTTGCCCAATGGGATGATTTCCACCAAGAGCTCA 294  
OY 1736 CCAGCATTTTCCACAGAGATGACAAATTTGAGCCCTTGGAGTTCCAGTGGATTCAGGA 1795  
|||||  
Db 293 CCAGCATTTTCCACAGAGATGACAAATTTGAGCCCTTGGAGTTCCAGTGGATTCAGGA 234  
OY 1796 AGGAGTGGGAAAGAGTTGGATGCTTATGAGCTTATGACATCAAGCTATGCTGATA 1855  
|||||  
Db 233 AGGAGTGGGAAAGAGTTGGATGCTTATGAGCTTATGACATCAAGCTATGCTGATA 174  
OY 1856 TCAGAAATATGAACAAATCTCTGCACAAAGAGAGAGCTTTAAGTTCCAGAGGTGCC 1915  
|||||  
Db 173 TCAGAAATATGAACAAATCTCTGCACAAAGAGAGAGCTTTAAGTTCCAGAGGTGCC 114  
OY 1916 TGGGCTCATTTGATATCTCTTCCCTCTGCAATTTTCCATACATAGAACCTTGAC 1975  
|||||  
Db 113 TGGGCTCATTTGATATCTCTTCCCTCTGCAATTTTCCATACATAGAACCTTGAC 54  
OY 1976 CTGCAAGCTGCATCTGTTAATCTAAATTTCCCAATTA 2016  
|||||  
Db 53 CTGCAAGCTGCATCTGTTAATCTAAATTTCCCAATTA 13

RESULT 3  
BM129080 571 bp mRNA linear EST 12-MAR-2002  
LOCUS  
DEFINITION f1f7c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
CDNA clone IMAGE:567655 5', similar to PR:Q9Y5R3 Q9Y5R3  
N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.;, mRNA sequence.  
ACCESSION BM129080  
VERSION BM129080.1 GI:17123632  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 571)  
Melton,D., Brown,J., Kenty,G., Permut,A., Lee,C., Kaestner,K.,  
Lemishka,I., Seearce,M., Brestelli,J., Gradowh,J., Clifton,S.,  
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Alstein,A.,  
Schmitt,A., Theising,B., Rutter,E., Ronko,I., Bennett,J., Cardenas  
M., Gibbons,M., McCann,R., Cole,R., Tsagaratshvili,R., Williams,T.,  
Jackson,Y. and Bowers,Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Other\_ESTS: f1f7c04.x1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biochem.harvard.edu  
Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center for information on  
obtaining a clone please contact: Juliana Brown  
(brown@fas.harvard.edu) This sequence now available from the IMAGE  
consortium, for clone orders contact: info@image.llnl.gov  
High quality sequence stop: 434.

FEATURES  
SOURCELocation/Qualifiers  
1. 571

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:567655"  
/clone\_lib="Melton Normalized Human Islet 4 N4-HIS 1"  
/sex="Both"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site\_1: Not 1; Site\_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 109 a 169 c 143 g 150 t  
ORIGIN

Query Match 27 1%; Score 551; DB 13; Length 571;  
Best Local Similarity 100.0%; Pred. No. 6,1e-162;  
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 TGAGTCTCAGCAGTGTAAAGCTTTACTCTTACAGCTTCTGGAGCGAGTGTCTTC 147  
Db 1 TGAGTCTCAGCAGTGTAAAGCTTTACTCTTACAGCTTCTGGAGCGAGTGTCTTC 60  
QY 148 AAGCCCGCTTGCAGGCTTCCACTCAGACATCTACTGCTTAAATAAAGCT 207  
Db 61 AAGCCCGCTTGCAGGCTTCCACTCAGACATCTACTGCTTAAATAAAGCT 120  
QY 208 CCTGCTTTCTGTTTCCAGATGAGCCATCTTGCTTCTTCCACATGTACAGCA 267  
Db 121 CCTGCTTTCTGTTTCCAGATGAGCCATCTTGCTTCTTCCACATGTACAGCA 180  
QY 268 CAACATCAGCTCCCTGTCTATGAAGCAGACCCGAGCCATGACCTGCTTCTGTC 327  
Db 181 CAACATCAGCTCCCTGTCTATGAAGCAGACCCGAGCCATGACCTGCTTCTGTC 240  
QY 328 TTCCGGGGCTCTGCTCTCTCTTTTGTGGGAGCTTTTGGGAGCACCAGATGTTT 387  
Db 241 TTCCGGGGCTCTGCTCTCTCTTTTGTGGGAGCTTTTGGGAGCACCAGATGTTT 300  
QY 388 CTACCTGATGAGCCGCTGCGACAGTGTGATGACCTTCAAGAGAGACCCGCTGAT 447  
Db 301 CTACCTGATGAGCCGCTGCGACAGTGTGATGACCTTCAAGAGAGACCCGCTGAT 360  
QY 448 GCTGCACATGGCTGTGGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGTCT 507  
Db 361 GCTGCACATGGCTGTGGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGTCT 420  
QY 508 TGATGCTTACATGGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAACAG 567  
Db 421 TGATGCTTACATGGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAACAG 480  
QY 568 CCGGGCCCTGTGTTCTGACCTGCTGTGACATCATCCACAAGATCAATCCCCCG 627  
Db 481 CCGGGCCCTGTGTTCTGACCTGCTGTGACATCATCCACAAGATCAATCCCCCG 540  
QY 628 GGCTCAGTCA 638  
Db 541 GGCTCAGTCA 551

RESULT 4

BM128370/c  
LOCUS BM128370 553 bp mRNA linear EST 12-MAR-2002  
DEFINITION I13d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
ACCESSION CDNA clone IMAGE:5676335 3', mRNA sequence.  
VERSION BM128370  
KEYWORDS BM128370.1 GI:17122822  
SOURCE EST.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 553)  
AUTHORS Melton,D., Brown,J., Kently,G., Penmutt,A., Lee,C., Kaestner,K., Lemishka,I., Scaerce,M., Brestcailly,J., Gradwohl,G., Clifton,S., Hillier,L., Maria,M., Pape,D., Wylie,T., Martin,J., Blistrain,A., Schmitt,A., Treising,B., Ralfer,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y., and Bowers,Y.  
TITLE Endocrine Pancreas Consortium  
JOURNAL Unpublished (2000)  
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownjfas.harvard.edu) This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov  
High quality sequence stop: 425.

LOCATION/Qualifiers  
1. 553

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5676335"  
/clone\_lib="Melton Normalized Human Islet 4 N4-HIS 1"  
/sex="Both"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site\_1: Not 1; Site\_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t  
ORIGIN

Query Match 23 5%; Score 478; DB 13; Length 553;  
Best Local Similarity 100.0%; Pred. No. 3.3e-139;  
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1540 TATGTGAGAGACATCCACAGTGAAGAGGATGCTCTTCTTTCTTGATCT 1599  
Db 553 TATGTGAGAGACATCCACAGTGAAGAGGATGCTCTTCTTTCTTGATCT 494  
QY 1600 TCCCTGTGGGCGAGACTTTCAGAGACTTGTGGCTTGAAGCCATTAAAGACAGACACT 1659  
Db 493 TCCCTGTGGGCGAGACTTTCAGAGACTTGTGGCTTGAAGCCATTAAAGACAGACAGT 434

QY 1660 ATCACTGATTTGATCCATAAACCCTCCTGTCCACATTTGCCCCAATGGGAATGATCT 1719  
 |||||||  
 Db 433 ATCACTGATTTGATCCATAAACCCTCCTGTCCACATTTGCCCCAATGGGAATGATCT 374  
 |||||||  
 QY 1720 TTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGCAATTTGAGACCTTGAGTTC 1779  
 |||||||  
 Db 373 TTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGCAATTTGAGACCTTGAGTTC 314  
 |||||||  
 QY 1780 CCAGTGGATTCAAGGAGGAAGTGGGAACAAGTTGATGCTTACTATAGACTTGACCA 1839  
 |||||||  
 Db 313 CCAGTGGATTCAAGGAGGAAGTGGGAACAAGTTGATGCTTACTATAGACTTGACCA 254  
 |||||||  
 QY 1840 TCACAGCATGGTATGAGAAATATGAACAATCTCTGCACAAAAGACCAACCTCT 1899  
 |||||||  
 Db 253 TCACAGCATGGTATGAGAAATATGAACAATCTCTGCACAAAAGACCAACCTCT 194  
 |||||||  
 QY 1900 AAGTTCACAGGAGTGGCTGGCTGATTTGATATACATTCCTCCCTCTGCATTTTCCATCA 1959  
 |||||||  
 Db 193 AAGTTCACAGGAGTGGCTGGCTGATTTGATATACATTCCTCCCTCTGCATTTTCCATCA 134  
 |||||||  
 QY 1960 CATGAGAACTTTGACCTGTGAGAGCTGCCATCTGTATATCTAAATTTCCCAATTAAG 2017  
 |||||||  
 Db 133 CATGAGAACTTTGACCTGTGAGAGCTGCCATCTGTATATCTAAATTTCCCAATTAAG 76  
 |||||||

## RESULT 5

BM128831/C

LOCUS 553 bp mRNA linear EST 12-MAR-2002

DEFINITION 117c04.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens

cDNA clone IMAGE:5676655 3', mRNA sequence.

ACCESSION BM128831

VERSION BM128831.1

KEYWORDS GI:17123383

SOURCE EST.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 553)

MELTON, D., BROWN, J., KENTY, G., PERMUTT, A., LEE, C., KAESTNER, K.,

LEHMSHA, I., SCAERCE, M., BRESTELL, J., GRADWOLL, G., CLIFTON, S.,

HILLIER, L., MARRA, M., PAPE, D., WYLIE, T., MARTIN, T., BLISTAIN, A.,

SCHMITT, A., THEISING, B., KILTER, E., KONKO, I., BENNETT, T., CARDENAS

, M., GIBBONS, M., MCCANN, R., COLE, R., TSAGAREISHVILI, R., WILLIAMS, T.,

JACKSON, Y. and BOWERS, Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, &amp; Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@ionp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center for information on

obtaining a clone please contact: Juliana Brown

(brown@fas.harvard.edu) This sequence now available from the IMAGE

consortium, for clone orders contact: info@image.llnl.gov

High quality sequence step: 425.

Location/Qualifiers

1..553

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5676655"

BASE COUNT 163 a 111 c 127 g 152 t

## ORIGIN

Query Match 23.5%; Score 478; DB 13; Length 553;

Best Local Similarity 100.0%; Pred. No. 3.3e-139;

Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 6

AL709927

LOCUS 668 bp mRNA linear EST 22-MAR-2002

DEFINITION DKEF268602364.F1 686 (synonym: h1cc3) Homo sapiens cDNA clone

DKEF268602364 5', mRNA sequence.

ACCESSION AL709927

VERSION AL709927.1

KEYWORDS GI:19693282

SOURCE EST.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 668)

DUESTERHOEFT, A., LAUBER, J., MEWES, H.W., GASENHUBER, J. and WIEMANN

, S.

EST (Duesterhoeft, et al.)

Unpublished (1999)

Contact: Duesterhoeft A

MIPS

Am Klopferstr. 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.

This clone (DKFZ68602364) is available at the RZPD in Berlin. Please contact the RZPD: Resourcezentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

## SOURCE

Location/Qualifiers

1..668

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZ68602364"

/clone\_lib="586 (synonym: hlc33)"

/tissue\_type="human skeletal muscle"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pTRIPlex2; Site\_1: sf11A; Site\_2: sf11B; cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others

## ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 468; DB 9; Length 668;

Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## FEATURES

## source

Location/Qualifiers

1..680

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="EFO109"

/dev\_stage="Adult"

/note="Organ: Lung-tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research)

profiles from the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 168 a 164 c 163 g 184 t 1 others

## ORIGIN

Query Match 15.6%; Score 318; DB 12; Length 680; Best Local Similarity 100.0%; Pred. No. 1.8e-89; Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
BF878439  
LOCUS 680 bp mRNA linear EST 17-JAN-2001  
DEFINITION MR0-EFO109-191100-002-h06 EFO109 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF878439  
VERSION BF878439.1 GI:12288569  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 680)  
Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Brines,M.R., Nagai,M.A., da Silva,M. Jr., Zago,M.S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brustein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL.  
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=1-MR0&t2=MR0-EFO109-191100-002-h06&t3=2000-11-19&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 17  
High quality sequence stop: 641.  
Location/Qualifiers  
1..680  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="EFO109"  
/dev\_stage="Adult"  
/note="Organ: Lung-tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles from the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 168 a 164 c 163 g 184 t 1 others  
ORIGIN  
Query Match 15.6%; Score 318; DB 12; Length 680; Best Local Similarity 100.0%; Pred. No. 1.8e-89; Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

VERSION      AR002418.1  GI:5849334
KEYWORDS     EST.
SOURCE       Homo sapiens
ORGANISM     human.
REFERENCE    Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT       Tumor Gene Index
              Unpublished (1997)
              Contact: Robert Strausberg, Ph.D.
              Email: cgaps-remail.nih.gov
              Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
              R. Emmert-Buck, M.D., Ph.D.
              CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
              Bernaldo, Ph.D.
              CDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/BLNt at:
              www.bio.linn.gov/bdnp/image/image.html
              Seq primer: -400p from G1pcc.
              Location/Qualifiers
                1..362
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_id="NCI-CGAP_GC6"
                /tissue_type="pooled germ cell tumors"
                /lab_host="DH10B"
                /note="Vector: pT73D-Pac (Pharmacia) with a modified
                polylinker. Site 1: Not I; Site 2: Eco RI; Plasmid DNA
                from the normalized library NCI-CGAP_GC4 was prepared, and
                ss circles were made in vitro. Following Hsp purification,
                this DNA was used as tracer in a subtractive hybridization
                reaction. The driver was PCR-amplified cDNAs from a pool
                of 5,000 clones made from the same library (cloneids
                1257096-1258631, 1469064-1470983, and 1475592-1476743).
                Subtraction by Bento Soares and M. Fatima Bernaldo. "
BASE COUNT   108 a 69 c 77 g 108 t
ORIGIN
Query Match 14.2%; Score 288; DB 10; Length 362;
Best Local Similarity 100.0%; Pred. No. 6.3e-80;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1730 AGCTACACAGCATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGTCCCAAGTGAATT 1789
DB 362 AGCTACACAGCATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGTCCCAAGTGAATT 303
QY 1790 CAAGGAGAGAGTGGGAGCAAGGTGGATGCTTACTTATGAGCTTGACCATCAGAGCTAT 1849
DB 302 CAAGGAGAGAGTGGGAGCAAGGTGGATGCTTACTTATGAGCTTGACCATCAGAGCTAT 243
QY 1850 CGGTATATAGAAATATGAAACAAATCTCTGACACAAAAGACAGCTCTTAAGTTCACAG 1909
DB 242 CGGTATATAGAAATATGAAACAAATCTCTGACACAAAAGACAGCTCTTAAGTTCACAG 183
QY 1910 GGTGCGCTGGGCTGCATTTGATATACACTTCCCTCTGCGATTTCCCATCAGATGAGAGAC 1969
DB 182 GGTGCGCTGGGCTGCATTTGATATACACTTCCCTCTGCGATTTCCCATCAGATGAGAGAC 123
QY 1970 TTGACCTGTGAAGTGCATCTGTTATTAATAAATTCCAATAAG 2017
DB 122 TTGACCTGTGAAGTGCATCTGTTATTAATAAATTCCAATAAG 75

```

```

RESULT 9
AM572390/c 419 bp mRNA linear EST 13-MAR-2000
LOCUS      x090a06.x2 NCI-CGAP_Col4 Homo sapiens cDNA clone IMAGE:2799634 3'
DEFINITION mRNA sequence.

```

```

ACCESSION    AM572390
VERSION      AM572390.1  GI:7237123
KEYWORDS     EST.
SOURCE       Homo sapiens
ORGANISM     human.
REFERENCE    Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT       Tumor Gene Index
              Unpublished (1997)
              Contact: Robert Strausberg, Ph.D.
              Email: cgaps-remail.nih.gov
              Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D.
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/BLNt at:
              image.linn.gov/image/html/resources.shtml
              Seq primer: -400p from G1pcc
              High quality sequence stop: 343.
              Location/Qualifiers
                1..419
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_id="IMAGE:2799634"
                /tissue_type="moderately-differentiated adenocarcinoma"
                /lab_host="DH10B"
                /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: 0190 dt.
                Average insert size 1.7 Kb. Life Technologies catalog #:
                11531-019"
BASE COUNT   122 a 78 c 89 g 130 t
ORIGIN
Query Match 13.9%; Score 283; DB 10; Length 419;
Best Local Similarity 99.7%; Pred. No. 2e-78;
Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1684 TCCCTGTCCACATCTTCCCAATGGGAGATGATCTTCCACCAAGAGCTCACAGCATT 1743
DB 419 TCCCTGTCCACATCTTCCCAATGGGAGATGATCTTCCACCAAGAGCTCACAGCATT 360
QY 1744 TTCCACAGAGATGCAATTTTGAGCCCTTGAGTCCCAAGTTCAGAGAGAGAGTG 1803
DB 359 TTCCACAGAGATGCAATTTTGAGCCCTTGAGTCCCAAGTTCAGAGAGAGAGTG 300
QY 1804 GGAACAAGTGTGATGCTACTTATGAGCTTGACCATCAGCTATCGTAATCAGAAAT 1863
DB 299 GGAACAAGTGTGATGCTACTTATGAGCTTGACCATCAGCTATCGTAATCAGAAAT 240
QY 1864 ATGAACAAATCTCTGACACAAAAGACAGCTCTTAAGTTCACAGAGGTCCTGCGCTGC 1923
DB 239 ATGAACAAATCTCTGACACAAAAGACAGCTCTTAAGTTCACAGAGGTCCTGCGCTGC 180
QY 1924 ATTGAATATCACCTTCCCTCTGCGATTTCCCATCAGATGAGAGCTTGAAGTGAAG 1983
DB 179 ATTGAATATCACCTTCCCTCTGCGATTTCCCATCAGATGAGAGCTTGAAGTGAAG 120
QY 1984 CTGCCATCTGTAAATACTAAATTCACCAATAAG 2017
DB 119 CTGCCATCTGTAAATACTAAATTCACCAATAAG 86

```

```

RESULT 10
AM572510/c 417 bp mRNA linear EST 13-MAR-2000
LOCUS      xq18g11.x2 NCI-CGAP_Ot1 Homo sapiens cDNA clone IMAGE:2751044 3'
DEFINITION mRNA sequence.

```

ACCESSION AM572510  
 VERSION AM572510.1 GI:7237243  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 417)  
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-rc@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov/image/html/resources.shtml](http://image.llnl.gov/image/html/resources.shtml)  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 416.  
 Location/Qualifiers  
 1..417  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2751044"  
 /clone\_lib="NCI-CCAP-Ut1"  
 /tissue\_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"  
 BASE COUNT 140 a 65 c 80 g 132 t  
 ORIGIN  
 Query Match 7.4%; Score 150; DB 10; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-37;  
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1868 AACAAATCTCTGCACAAAGAGCAGCTTAAAGTTCACAGGCGTGGCGTGCATTT 1927  
 DB 383 AACAAATCTCTGCACAAAGAGCAGCTTAAAGTTCACAGGCGTGGCGTGCATTT 324  
 QY 1928 GAATATCACTTCCCTCTGCATTTCCCATCACAATAGAACTTGAACCTGTGAAGCTGC 1967  
 DB 323 GAATATCACTTCCCTCTGCATTTCCCATCACAATAGAACTTGAACCTGTGAAGCTGC 264  
 QY 1988 CATCTGTAATACATAAATTCACCAATTAAG 2017  
 DB 263 CATCTGTAATACATAAATTCACCAATTAAG 234  
 RESULT 11  
 BE056840/c 358 bp mRNA linear EST 16-OCT-2000  
 LOCUS 7x10d12.x1 NCI-CCAP-GC6 Homo sapiens cDNA clone IMAGE:3443783.3  
 DEFINITION mRNA sequence.  
 ACCESSION BE056840  
 VERSION BE056840.1 GI:10810736  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 358)  
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-rc@mail.nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: [info@image.llnl.gov](mailto:info@image.llnl.gov)  
 Seq primer: -400P from Gibco.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..358  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3443783"  
 /clone\_lib="NCI-CCAP-GC6"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI-CCAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1459064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 107 a 68 c 82 g 101 t  
 ORIGIN  
 Query Match 6.7%; Score 136; DB 12; Length 358;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-32;  
 Matches 136; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
 QY 1730 AGCTCACAGCATTTTCCACAGATGATGCAATTCGACCCCTTGAGTCCAGTGGATT 1789  
 DB 358 AGCTCACAGCATTTTCCACAGATGATGCAATTCGACCCCTTGAGTCCAGTGGATT 299  
 QY 1790 CANGAAGAGAGTGGACACAGGTTGGATGCTTATGAGCTTGACCATCAGCTAT 1849  
 DB 298 CANGAAGAGAGTGGACACAGGTTGGATGCTTATGAGCTTGACCATCAGCTAT 239  
 QY 1850 CGGTAATCAGAAATAT 1865  
 DB 238 CGGTAATCAGAAATAT 223  
 RESULT 12  
 BF197521/c 525 bp mRNA linear EST 03-NOV-2000  
 LOCUS 7084a08.x1 NCI-CCAP\_K1d11 Homo sapiens cDNA clone IMAGE:3642903.3  
 DEFINITION similar to TR:Q90BY3 Q90BY3 N-ACETYLBUCOSAMINE 6-O-SULFOTRANSFERASE.; mRNA sequence.  
 ACCESSION BF197521  
 VERSION BF197521.1 GI:11086670  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 525)  
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-rc@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

FEATURES  
source  
Location/Qualifiers  
1..525

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3642903"  
/clone\_lib="NCI-CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT  
ORIGIN  
90 a 163 c 187 g 85 t

Query Match 2.9%; Score 59; DB 12; Length 525;  
Best Local Similarity 100.0%; Pred. No. 9,1e-09;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 736  
|||||  
Db 260 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 202

RESULT 13  
A1824100/c  
LOCUS  
DEFINITION  
A1824100 620 bp mRNA linear EST 21-DEC-1999  
w146c01.x1 NCI-CGAP\_Lu19 Homo sapiens CDNA clone IMAGE:2405856 3'  
similar to TR:075667 075667 DJ71116.4; contains PTRS.b2 PTR5  
repetitive element; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 620)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrrp/image/image.html  
Insert Length: 917 Std Error: 0.00  
Seq primer: -40UP from G1bco  
High quality sequence stop: 490.

FEATURES  
source

Location/Qualifiers  
1..620  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2405856"  
/clone\_lib="NCI-CGAP\_Lu19"  
/tissue\_type="squamous cell carcinoma, poorly

differentiated (4 pooled tumors, including primary and  
metastatic)"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
pooled lung tumor tissue, and was then primed with a Not I  
- oligo(dT) primer. Double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT7T3 vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT  
ORIGIN  
106 a 195 c 227 g 89 t 3 others

Query Match 2.9%; Score 59; DB 9; Length 620;  
Best Local Similarity 100.0%; Pred. No. 8e-09;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 736  
|||||  
Db 314 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 256

RESULT 14  
BE857538/c  
LOCUS  
DEFINITION  
BE857538 695 bp mRNA linear EST 29-SEP-2000  
7901a08.x1 NCI-CGAP\_Brn23 Homo sapiens CDNA clone IMAGE:3305174 3'  
similar to TR:09UBY3 09UBY3 N-ACETYLGLUCOSAMINE  
6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 695)  
NCI/NIHNS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BRGAP), Tumor Gene Index  
Unpublished (1998)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from G1bco  
High quality sequence stop: 431.

FEATURES  
source

Location/Qualifiers  
1..695  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3305174"  
/clone\_lib="NCI-CGAP\_Brn23"  
/tissue\_type="glioblastoma (pooled)"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTT  
T 3'); double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and Eco RI  
library is normalized, and was constructed by Bento

BASE COUNT 127 a 210 c 247 g 111 t  
 ORIGIN  
 Query Match 2.9%; Score 59; DB 12; Length 695;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-09;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTCTTCAACCTGCAG 736  
 |||||||  
 Db 272 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTCTTCAACCTGCAG 214

RESULT 15  
 BE858652/c 735 bp mRNA linear EST 29-SEP-2000  
 LOCUS 7401a09.x1 NCI-CGAP-Brn23 Homo sapiens cDNA clone IMAGE:3305176 3'  
 DEFINITION Similar to TR:09UBY3 09UBY3 N-ACETYLGLUCOSAMINE  
 6-O-SULFOTRANSFERASE.; mRNA sequence.  
 ACCESSION BE858652  
 VERSION BE858652.1 GI:103738930  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 735)  
 NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/BRGAP), Tumor Gene Index  
 Unpublished (1998)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgepbs.r@mail.nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D. Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 425.  
 Location/Qualifiers  
 1..735  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3305176"  
 /clone\_lib="NCI-CGAP-Brn23"  
 /tissue\_type="glioblastoma (pooled)"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pTZ193-Pac (Pharmacia) with a  
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGAAGTGGAGGCGCCGACATCTTTTCTTTTCTTTTCTTTT  
 T 3'] double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pTZ193 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 221 c 253 g 120 t 6 others  
 ORIGIN  
 Query Match 2.9%; Score 59; DB 12; Length 735;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-09;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTCTTCAACCTGCAG 736  
 |||||||  
 Db 328 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTCTTCAACCTGCAG 270

RESULT 16  
 AM081348/c 436 bp mRNA linear EST 14-OCT-1999  
 LOCUS xc41b06.x1 NCI-CGAP Co20 Homo sapiens cDNA clone IMAGE:2586803 3'  
 DEFINITION similar to TR:088195 088199 CHONDROITIN 6-SULFOTRANSFERASE.; mRNA  
 sequence.  
 ACCESSION AM081348  
 VERSION AM081348.1 GI:60365500  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 436)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgepbs.r@mail.nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
 Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The  
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
 Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bdtp/image/image.html  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 431.  
 Location/Qualifiers  
 1..436  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2586803"  
 /clone\_lib="NCI-CGAP-Co20"  
 /tissue\_type="moderately differentiated adenocarcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site.1: SalI;  
 Site.2: NotI; Cloned unidirectionally. Primer: oligo dr.  
 Normalized to cot 500. Average insert size 1.1kb.  
 Normalized version of NCI-CGAP-Co18. Library constructed  
 by Life Technologies."

BASE COUNT 64 a 136 c 169 g 67 t  
 ORIGIN  
 Query Match 2.3%; Score 47; DB 10; Length 436;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTC 724  
 |||||||  
 Db 309 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTC 263

RESULT 17  
 BM246681 497 bp mRNA linear EST 01-FEB-2002  
 LOCUS K0741E04-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit-/sca-1-)  
 DEFINITION CDNA Library (Long) Mus musculus cDNA clone K0741E04 3', mRNA  
 sequence.  
 ACCESSION BM246681  
 VERSION BM246681.1 GI:17881951  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 497)  
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,  
 Alba, K., Taub, D., Longo, D.D., Keller, J. and Ko, M.S.H.  
 Systematic Analyses of NIA Mouse Hematopoietic Stem Cell



JOURNAL  
COMMENT

(lin/-c-Klt-/Sca-1-) cDNA library (Long)  
 Unpublished (2001)  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 Plate: K0741 row: E column: 04  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 497  
 POLYA=yes

## FEATURES

## SOURCE

Location/Qualifiers  
 1..497  
 /organism="Mus musculus"  
 /strain="C57BL/6NCr"  
 /db\_xref="taeST:K0741E04-3"  
 /db\_xref="taxon:10090"  
 /clone="K0741E04"  
 /clone\_lib="NIA Mouse Hematopoietic Stem Cell  
 (lin/-c-Klt-/Sca-1-) cDNA library (Long)"  
 /tissue\_type="Hematopoietic Stem Cell (lin/-c-Klt-/Sca-1-)"  
 /dev\_stage="Age approx.10 weeks old"  
 /lab\_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Cancer Institute on Aging, USA), Jonathan Keller (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCGATCGCGAGCGCCGCTTTT-3'] from 0.9 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.1 Kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT  
ORIGIN

87 a 137 c 134 g 139 t

## Query Match

1.3%; Score 27; DB 13; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 494 GACATGACGCTTTGATGCTACATG 520  
 ||||||||||||||||||  
 Db 441 GACATGAGCGCTTTGATGCTACATG 467

## RESULT 18

## BM245312

## LOCUS

## DEFINITION

BM245312 536 bp mRNA linear EST 01-FEB-2002  
 K0722H07-3 NIA Mouse Hematopoietic Stem Cell (lin/-c-Klt-/Sca-1-)  
 cDNA library (Long) Mus musculus cDNA clone K0722H07 3', mRNA  
 sequence.

ACCESSION BM245312.1 GI:17880582

## VERSION

## KEYWORDS

## SOURCE

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 536)  
 Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G.,  
 Alba,K., Taub,D., Longo,D.L., Keller,J. and Ko,M.S.H.  
 Systematic Analyses of NIA Mouse Hematopoietic Stem Cell  
 (lin/-c-Klt-/Sca-1-) cDNA Library (Long)

## TITLE

## JOURNAL

## COMMENT

Unpublished (2001)  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 Plate: K0722 row: H column: 07  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 536  
 POLYA=yes

## FEATURES

## SOURCE

Location/Qualifiers  
 1..536  
 /organism="Mus musculus"  
 /strain="C57BL/6NCr"  
 /db\_xref="taeST:K0722H07-3"  
 /db\_xref="taxon:10090"  
 /clone="K0722H07"  
 /clone\_lib="NIA Mouse Hematopoietic Stem Cell  
 (lin/-c-Klt-/Sca-1-) cDNA library (Long)"  
 /tissue\_type="Hematopoietic Stem Cell (lin/-c-Klt-/Sca-1-)"  
 /dev\_stage="Age approx.10 weeks old"  
 /lab\_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Cancer Institute on Aging, USA), Jonathan Keller (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCGATCGCGAGCGCCGCTTTT-3'] from 0.9 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.1 Kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT  
ORIGIN

93 a 150 c 147 g 146 t

## Query Match

1.3%; Score 27; DB 13; Length 536;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 494 GACATGACGCTTTGATGCTACATG 520  
 ||||||||||||||||||  
 Db 441 GACATGAGCGCTTTGATGCTACATG 467

## RESULT 19

## BG964671

## LOCUS

## DEFINITION

BG964671 695 bp mRNA linear EST 12-JUN-2001  
 602831875F1 NCI-CGAP\_C024 Mus musculus cDNA clone IMAGE:4966258 5',  
 mRNA sequence.

ACCESSION BG964671.1 GI:14352308

## VERSION

## KEYWORDS

## SOURCE

house mouse.

```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          NIH-MGC http://mgc.nci.nih.gov/
JOURNAL        National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT        Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                CDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLM10985 row: f column: 19
                High quality sequence stop: 684.
                Location/Qualifiers
                1..695
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4982643"
                /clone_1lb="NCI_CGAP_Co24"
                /lab_host="DH10B (TI phage-resistant)"
                /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
                Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.6 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT    118 a      216 c      195 g      166 t

ORIGIN
Query Match      1.3%; Score 27; DB 13; Length 695;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 494 GACATGAGCGCTTTGATGCTCCTACATG 520
|||||
DB 345 GACATGAGCGCTTTGATGCTCCTACATG 371

RESULT 20
LOCUS          BG963298 783 bp mRNA linear EST 12-JUN-2001
DEFINITION    60282716F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
ACCESSION     BG963298
VERSION       BG963298
KEYWORDS      EST
SOURCE        house mouse.
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          NIH-MGC http://mgc.nci.nih.gov/
JOURNAL        National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT        Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                CDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLM10985 row: p column: 04
                High quality sequence stop: 741.
                Location/Qualifiers
                1..783
                /organism="Mus musculus"
                /strain="FVB/N"

FEATURES
Source
Query Match      1.3%; Score 27; DB 13; Length 852;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 494 GACATGAGCGCTTTGATGCTCCTACATG 520
|||||
DB 447 GACATGAGCGCTTTGATGCTCCTACATG 473

RESULT 21
LOCUS          BG966340 952 bp mRNA linear EST 12-JUN-2001
DEFINITION    602832826F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4987327 5',
ACCESSION     BG966340
VERSION       BG966340
KEYWORDS      EST
SOURCE        house mouse.
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          NIH-MGC http://mgc.nci.nih.gov/
JOURNAL        National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT        Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                CDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLM10998 row: c column: 08
                High quality sequence stop: 784.
                Location/Qualifiers
                1..852
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4987327"
                /clone_1lb="NCI_CGAP_Co24"
                /lab_host="DH10B (TI phage-resistant)"
                /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
                Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.6 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT    176 a      248 c      244 g      184 t

ORIGIN
Query Match      1.3%; Score 27; DB 13; Length 852;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 494 GACATGAGCGCTTTGATGCTCCTACATG 520
|||||
DB 447 GACATGAGCGCTTTGATGCTCCTACATG 473

RESULT 22
LOCUS          AK009113
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

LOCUS	1923 bp	LINEAR	HTC 19-JAN-2000
DEFINITION	Mus musculus adult male tongue CDNA, RIKEN full-length enriched library, clone:2310003618:carbohydrate (chondroitin 6/sulfate)		
ACCESSION	AK009113		
VERSION	AK009113.1 GI:12843701		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male tongue CDNA to mRNA, clone:lib:RIKEN full-length enriched mouse CDNA library clone:2310003618.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishi, Y., Nakamura, S., Hazama, N., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubuchi, Y., Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishi, Y., Aizawa, T., Hara, A., Piskunish, Y., Kono, H., Adachi, J., Fukuda, S., Aikawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojocori, T., Bono, H., Kasukawa, T., Saito, R., Kado, K., Matsuda, H., Ashburner, M., Batalov, S., Cassavani, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, T., Pesole, G., Quackenbush, J., Schriml, L. M., Stambli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Baren, G., Blake, J., Boffelli, D., Bojunga, N., Cerinchi, P., de Bonaldo, M. F., Brownstein, M. J., But, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotman, M., Hume, D. A., Kamuya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, J., Sakemoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-oka, K., Wang, K. H., Wetz, C., Whitaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, I., Kawaji, H., Kotsuki, S. and Hayashizaki, Y.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409 (6821), 685-690 (2001)		
MEDLINE	21085660		
PUBMED	11217851		
REFERENCE	5 (bases 1 to 1923)		
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arikawa, T., Balderelli, R., Bono, H., Brownstein, M., But, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imocanti, K., Ishi, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,		

TITLE  
JOURNAL

Saito,H., Saito,R., Sakai,K., Sakai,K., Sano,H., Sasaki,D.,  
Schirml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,  
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and  
Hayashizaki,Y.

Direct Submission  
Submitted (10-Jul-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp).  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGAGAAGACGATCCAGACCGCTTTTTCCTTTTTTTTN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 5.0 and subtraction to  
Rot = 25.0. Second strand cDNA was prepared with the primer adapter  
of sequence 15',  
GAGAGAGAGATTCTCAGTTAATTAAATTATGCCCCCCCCC 3'. cDNA was cleaved  
with XhoI and SstI. Cloning sites, 5' end: XhoI, 3' end: SstI.  
Host: SOLR.

Location/Qualifiers

1..1923

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="FANTOM.DB:2310003G18"  
/cb\_xref="MGD:MGI:1904558"  
/db\_xref="taxon:10090"  
/clone="2310003G18"  
/sex="male"  
/tissue\_type="longue"  
/clone\_id="RIKEN.full-length enriched mouse cDNA library"  
/gen\_stage="adult"

1..1923

/gene="Chet4"  
96..1262  
/gene="Chet4"  
/note="carbohydrate (chondroitin 6/ketaran)  
sulfoltransferase 4  
data source:MGD, source key:MGI:1349479, evidence:ISS  
putative"

/codon\_start=1  
/protein\_id="BAB26078.1"  
/cd\_xref="GI:12833702"  
/cb\_xref="MGD:MGI:1349479"  
/translation="IMLLKKGRLIMFLFGSOIVVALLIHMSVRHLSORESPRPVH  
LVSSMGGSGSFVGOLFGRPDGVYLMEPRMHVMPTSTAKMLHAAYDDLRSVL  
CDMKEPKAVNMPGRKOSSLFOWQSQALSLAPRCDFPRAHEISSPKRHCLLCQQDF  
DMFKACRSHGFVYLKVEKLSDALPLLDLSLNLHVHLVAHPADPAKSRHHITL  
ELAWDSHILVGQHLLETIKEDQPYAAKIICKISOVDIVKAIIQLPEALQRYFLRE  
DIVAPLAQTTRLYKFVGLDFLPILQTVNVYNVARGKMGQHASHRTARNRLANVSQMR  
WSLPERKYVSLDDPCGEAMDLGLGYRSGOOGGNLSLDLSSHLIGVFREG"

BASE COUNT 445 a 533 c 492 g 453 t

Query Match 1.3% Score 27; DB 11; Length 1923;  
Best Local Similarity 100.0%; Prid. No. 31;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

494 GACATGACGCCTTGATGACGCAACTG 520  
|||||  
405 GACATGACGCCTTGATGACCTACTG 431

RESULT 23  
BM036473/c 501 bp mRNA linear EST 05-NOV-2001  
DEFINITION fufi07.x1 Gong zebrafish testis Danio rerio cDNA clone 5377309 3',  
ACCESSION BM036473  
VERSION BM036473.1 GI:16750044  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
; Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 501)  
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy  
S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,  
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
and Wilson,R.  
TITLE Washu Zebrafish EST Project 1998  
JOURNAL Unpublished (1998)  
COMMENT Contact: Stephen L. Johnson  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zdrafish@watson.wustl.edu  
The library was constructed by Dr. Z. Gong. DNA sequencing by:  
Washington University Genome Sequencing Center St. Louis. Please  
contact Zhiyuan Gong for further information on this library  
(National University of Singapore, Department of Biological  
Sciences, Lower Kent Ridge Road, Singapore 119260).  
High quality sequence stop: 367.  
Location/Qualifiers  
1. 501  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="5377309"  
/clone\_lib="Gong zebrafish testis"  
/sex="male"  
/dev\_stage="4-5 month"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis (pooled); Vector: pBluescript SK-  
Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from  
the testes of 31 male adult zebrafish (4-5 month old)  
cDNAs were made using oligo-dT primers and inserted into  
lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo  
mass-excisio to pBluescript SK- following the Washington  
University protocol  
(http://genome.wustl.edu/est/lambda\_protocol.shtml).  
Please contact Zhiyuan Gong for further information on  
this library (National University of Singapore,  
Department of Biological Sciences, Lower Kent Ridge Road,  
Singapore 119260)."  
BASE COUNT 171 a 94 c 90 g 146 t  
ORIGIN  
Query Match 1.2%; Score 25; DB 13; Length 501;  
Best Local Similarity 100.0%; Pred. No. 3 6e+02;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2008 CCCCAATAGAGAAAAA 2032  
Db 29 CCCCAATAGAGAAAAA 5  
RESULT 24.  
CNS02BK0/c 682 bp DNA linear GSS 12-MAY-2000  
LOCUS Tetradon nigroviridis genome survey sequence PCC-0r1 end of clone  
DEFINITION 253N23 of library G from Tetradon nigroviridis, genomic survey  
sequence.

ACCESSION AL189971  
VERSION AL189971.1 GI:7828075  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetradon nigroviridis.  
ORGANISM Tetradon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
REFERENCE 1 (bases 1 to 682)  
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetradon nigroviridis DNA sequence  
Unpublished  
2 (bases 1 to 682)  
REFERENCE Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradon nigroviridis  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 682)  
AUTHORS genome scope.  
JOURNAL Direct Submission  
COMMENT Submitted (12-APR-2000)  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetradon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetradon.  
Location/Qualifiers  
1. 682  
/organism="Tetradon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="253N23"  
/clone\_lib="G"  
/note="Genoscope sequence ID : CGAG253CG12SP1-end ;  
PCC-0r1"  
BASE COUNT 202 a 122 c 113 g 224 t 21 others  
ORIGIN  
Query Match 1.2%; Score 25; DB 17; Length 682;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2008 CCCCAATAGAGAAAAA 2032  
Db 241 CCCCAATAGAGAAAAA 217  
RESULT 25  
LOCUS BM663231/c 375 bp mRNA linear EST 27-FEB-2002  
DEFINITION UT-E-C10-aah-e-08-0-01.s1 UT-E-C10 Homo sapiens cDNA clone  
ACCESSION BM663231  
VERSION BM663231.1 GI:18968394  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 375)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 The following repetitive elements were found in this CDNA sequence: 1-42, >PCF1\_A#Simple\_repeat (matched compliment)  
 Seq primer: M13 Forward  
 POLYA-Yes

# FEATURES

## SOURCE

Location/Qualifiers  
 1..375  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-C10-ab-e-08-0-UI"  
 /clone\_lib="UI-E-C10"  
 /tissue\_type="RPE and Choroid"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I;  
 UI-E-C10 is a CDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
 TAG\_LIB=UI-E-C10  
 TAG\_TISSUE=RPE and Choroid  
 TAG\_SEQ=ACCTA"

BASE COUNT 88 a 82 c 89 g 116 t  
 ORIGIN

Query Match 1.2%; Score 24; DB 13; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 9.2e+02;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATAGAGAAAAA 2032  
 Db 26 CCAATAGAGAAAAA 3

RESULT 26  
 BM666130/c 566 bp mRNA linear EST 27-FEB-2002

LOCUS UI-E-DX1-agt-f-19-0-UI.s1 UI-E-DX1 Homo sapiens CDNA clone  
 DEFINITION UI-E-DX1-agt-f-19-0-UI 3', mRNA sequence.  
 ACCESION BM666130  
 VERSION BM666130.1 GI:18973587  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 566)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 The following repetitive elements were found in this CDNA sequence: 23-556, >11#1ME/L1 (matched compliment)  
 Seq primer: M13 Forward  
 POLYA-Yes.

# FEATURES

## SOURCE

Location/Qualifiers  
 1..566  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DX1-agt-f-19-0-UI"  
 /clone\_lib="UI-E-DX1"  
 /tissue\_type="fetal eyes"  
 /dev\_stage="fetal"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I;  
 UI-E-DX1 is a normalized CDNA library containing the following tissue(s): fetal eyes. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACATCAAGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
 TAG\_LIB=UI-E-DX1  
 TAG\_TISSUE=human fetal eyes  
 TAG\_SEQ=ACATCAAGA"

BASE COUNT 108 a 94 c 89 g 275 t  
 ORIGIN

Query Match 1.2%; Score 24; DB 13; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 6.8e+02;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATAGAGAAAAA 2032  
 Db 24 CCAATAGAGAAAAA 1

RESULT 27  
 CNS0127Y 919 bp DNA linear GSS 12-MAY-2000

LOCUS CNS0127Y  
 DEFINITION Tetraodon nigroviridis genome survey sequence pUC-ori end of clone 219308 of library G from Tetraodon nigroviridis, genomic survey sequence.  
 ACCESION AL173959  
 VERSION AL173959.1 GI:7812016  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis  
 Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
 REFERENCE 1 (bases 1 to 919)  
 Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzmes,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa

```

TITLE      Human gene number estimate provided by genome wide analysis using
JOURNAL    Tetradon nigroviridis DNA sequence
REFERENCE   Unpublished
AUTHORS     2 (bases 1 to 919)
            Rouest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Querier,F., Saurin,W., Bernot,A. and
            Weissensbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
JOURNAL    freshwater pufferfish Tetradon nigroviridis
REFERENCE   Unpublished
AUTHORS     3 (bases 1 to 919)
JOURNAL    Direct Submission
TITLE      Genoscope.
COMMENT     Submitted (12-APR-2000)
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetradon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/tetradon.
FEATURES    Location/Qualifiers
            1..919
            /organism="Tetradon nigroviridis"
            /db_xref="taxon:99863"
            /clone="219308"
            /clone_1lb="g"
            /note="Genoscope sequence ID : COAG219DE04SP1-end :
            PUC-ori"
BASE COUNT  314 a    151 c    164 g    283 t    7 others
ORIGIN
Query Match      1.2%: Score 24; DB 17; Length 919;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  2009 CCAATTAGAGAAAAAAGAAAAA 2032
      |||||||||||||||||||||||||
Db   514 CCAATTAGAGAAAAAAGAAAAA 537

RESULT 28
B75634/c      132 bp      DNA      linear      GSS 08-APR-1999
LOCUS      RPCI11-11M14.TV RPCI-11 Homo sapiens genomic clone RPCI-11-11M14,
DEFINITION  DNA sequence.
ACCESSION   B75634
VERSION     B75634.1 GI:2771321
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 132)
            Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
            ,J.C.
            Use of BAC End Sequences for Sequence-Ready Map Building
JOURNAL    Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-11M14.TP
            Contact: Marx Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
            Location/Qualifiers

```

```

source      1..132
            /organism="Homo sapiens"
            /db_xref="GDB:7504141"
            /db_xref="taxon:9606"
            /clone="RPCI-11-11M14"
            /clone_1lb="RPCI-11"
            /sex="Male"
            /cell_type="Lymphocytes"
            /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
            RPCI11 Human Male BAC Library"
BASE COUNT  36 a    19 c    20 g    57 t
ORIGIN
Query Match      1.1%: Score 23; DB 17; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.2e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  2010 CAAATAGAGAAAAAAGAAAAA 2032
      |||||||||||||||||||||||||
Db   107 CAAATAGAGAAAAAAGAAAAA 85

RESULT 29
A1072478/c    191 bp      mRNA      linear      EST 05-JUL-1999
LOCUS      UI-R-C2-nm-c-03-0-UI s1 UI-R-C2 Rattus norvegicus cDNA clone
DEFINITION  UI-R-C2-nm-c-03-0-UI 3', mRNA sequence.
ACCESSION   A1072478
VERSION     A1072478.1 GI:3398672
KEYWORDS    EST.
SOURCE      Norway rat.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 191)
REFERENCE   Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
COMMENT     97044477
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: mscoares@iue.uiowa.edu
            cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
            distribution: clones will be available through Research Genetics
            This clone is also available through the I.M.A.G.E. Consortium at
            LNL (info@image.llnl.gov). IMAGE ID=1786676
            Seg primer: M13 Forward
            POLYA-No.
FEATURES    Location/Qualifiers
            1..191
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-C2-nm-c-03-0-UI"
            /clone_1lb="UI-R-C2"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies)"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-C2
            library is a subtracted library derived from the UI-R-C1
            library, which is a subtracted library derived from the
            UI-R-C0 library. The UI-R-C0 library consisted of a
            mixture of individually tagged normalized libraries
            constructed from rat placenta, adult lung, brain, liver,
            kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
            embryo. The tag is a string of 3-5 nucleotides present
            between the Not I site and the oligo-dT track which allows

```

identification of the library of origin of a clone within the mixture. The subcloned library (UI-R-C2) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subcloned library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10b bacteria (Life Technologies) to generate the UI-R-C2 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996)".

BASE COUNT 47 a 41 c 30 g 73 t  
ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 191;  
Best Local Similarity 100.0%; Pred. No. 3.1e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032  
Db 25 CAATAGAGAAAAA 3

RESULT 30  
B0128285/c 269 bp mRNA linear EST 28-JUN-2000  
LOCUS B0128285  
DEFINITION Musculus cDNA clone 9630029B21 3', mRNA sequence.  
ACCESSION B0128285  
VERSION B0128285.1 GI:8782652  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.

## REFERENCE

AUTHORS Eukaryota: Metazoa; Chordata: Craniata; Vertebrata: Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 269)  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Konno, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takehashi, F., Tomiyaga, N., Toyota, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamazaki, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp/  
Carninci, P. and Hayashizaki, Y.

## TITLE

## JOURNAL

## COMMENT

Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

## FEATURES

## source

19-44 (1999)  
Please visit our web site (<http://genome.irc.riken.go.jp>) for further details.  
Location/Qualifiers

1..269  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="9630029B21"  
/clone\_11b="RIKEN full-length enriched, 16 days neonate cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="16 days neonate"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGACAGAGAGATCCAGACCTCTTTTCTTTTCTTNN 3']. cDNA was prepared by using triazole thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGACAGAGATCCAGATTAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified phuscript KS(+) after bulk excision from Lambda FVC I."

BASE COUNT 79 a 52 c 30 g 108 t  
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 269;  
Best Local Similarity 100.0%; Pred. No. 2.4e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032  
Db 159 CAATAGAGAAAAA 137

RESULT 31  
B0191911/c 304 bp mRNA linear EST 30-APR-2002  
LOCUS B0191911  
DEFINITION UI-R-DRI-CKY-g-20-0-UI-s1 UI-R-DRI Rattus norvegicus cDNA clone  
UI-R-DRI-CKY-g-20-0-UI 3', mRNA sequence.  
ACCESSION B0191911  
VERSION B0191911.1 GI:20367462  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 304)  
Bonaldi, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the

normalized osteoblast library cDNA Library Preparation: M.B. Soares  
Lab Clone distribution: clones will be available through Research  
Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-yes.

#### FEATURES

Source

Location/Qualifiers  
1. 304  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DRI-cky-g-20-0-UI"  
/clone\_lib="UI-R-DRI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DRI  
library is a normalized Rat Osteoblast library (nREO)  
constructed in pT73 vector according to the procedure  
described by Bonaldo, Lennon & Soares (Normalization and  
Subtraction: Two Approaches to Facilitate Gene Discovery,  
Genome Research 6: 791-806, 1996). The oligonucleotide  
used to prime first strand synthesis contained the  
sequence tag AACATATCAA between the Not I cloning site and  
dri8 stretch. The Rat Osteoblast tissue was provided by  
Lian & Stein of the University of Massachusetts Medical  
School.

TAG\_LIB=UI-R-DRI  
TAG\_TISSUE=osteoblast  
TAG\_SEQ=AACATATCAA

BASE COUNT 84 a 67 c 49 g 104 t  
ORIGIN

Query Match 1.1%; Score 23; DB 14; Length 304;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CCAATAGAAAAA 2032  
DB 23 CCAATAGAAAAA 1

RESULT 32  
BE117198/c 319 bp mRNA linear EST 13-JUN-2000  
LOCUS BE117198  
DEFINITION UI-R-BS1-ayh-a-02-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone  
UI-R-BS1-ayh-a-02-0-UI.3', mRNA sequence.  
BE117198  
VERSION BE117198.1 GI:8509303  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 319)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE  
COMMENT 97044477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.iowa.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
clones will be available through Research Genetics (www.resgen.com)  
The following repetitive elements were found in this cDNA sequence:

1-48, >POLY\_A#Simple-repeat  
Seq primer: M13 Forward  
POLYA-yes.

#### FEATURES

Source

Location/Qualifiers  
1. 319  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BS1-ayh-a-02-0-UI"  
/clone\_lib="UI-R-BS1"  
/dev\_stage="embryonic 13 dpc"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BS1  
library is derived from 13 dpc whole embryo tissue. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
ratest.eng.uiowa.edu.  
TAG\_SEQ=None found"

BASE COUNT 67 a 63 c 61 g 128 t  
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 319;  
Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATAGAAAAA 2031  
DB 23 CCAATAGAAAAA 1

RESULT 33  
BG291580 342 bp mRNA linear EST 21-FEB-2001  
LOCUS BG291580  
DEFINITION 60238567P1 NIH-MGC\_93 Homo sapiens cDNA clone IMAGE:4514872 5',  
mRNA sequence.  
BG291580  
VERSION BG291580.1 GI:13049825  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 342)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs.femail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LHAM10403 row: m column: 17  
High quality sequence stop: 342.

#### FEATURES

Source

Location/Qualifiers  
1. 342  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4514872"  
/clone\_lib="NIH-MGC\_93"  
/tissue\_type="transitional cell papilloma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library."

BASE COUNT 136 a 49 c 59 g 98 t  
ORIGIN



Query Match 1.1%; Score 23; DB 12; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 2e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032  
 ||||||||||||||||  
 Db 294 CAATATAGAAAAA 316

RESULT 34  
 B1555949 349 bp mRNA linear EST 05-SEP-2001  
 LOCUS 603239076F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5291971 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1555949  
 VERSION B1555949.1 GI:15443263  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 349)  
 NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue procurement: lothar Hennighausen Ph.D., Chu Xia Deng Ph.D.  
 CDNA library preparation: Life Technologies, Inc.  
 CDNA library arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 plate: LLM11738 row: h column: 20  
 High quality sequence stop: 349.  
 Location/Qualifiers  
 source  
 1.349  
 /organism="Mus musculus"  
 /strain="129,C57BL/6J,FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5291971"  
 /clone\_lib="NCI\_CGAP\_Mam3"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="10 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;  
 Site:2: NotI. Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Chu Xia Deng, NIH  
 Reference for transgenic model: Xu et al., Nature Genetics  
 22, 37-43 (1999)."

BASE COUNT 108 a 80 c 66 g 95 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 13; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 2e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032  
 ||||||||||||||||  
 Db 327 CAATATAGAAAAA 349

RESULT 35  
 BE340900 362 bp mRNA linear EST 17-JUL-2000  
 LOCUS EST344988 potato stolon, Cornell University Solanum tuberosum cDNA  
 DEFINITION clone cSNAL1K13, mRNA sequence.  
 ACCESSION BE340900  
 VERSION BE340900.1 GI:9250431  
 KEYWORDS EST.

SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 1 (bases 1 to 362)  
 van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,  
 Holt,I.E., Liang,F., Hansen,T.S., Utechtack,T., Bowman,C.L., Doan  
 B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker  
 R.  
 Generation of ESTs from potato swelling stolons  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Research Genetics, Libraries Division  
 Tel: 1-800-711-6195  
 Email: cdna@resgen.com  
 5 prime sequence.  
 Location/Qualifiers  
 source  
 1.362  
 /organism="Solanum tuberosum"  
 /cultivar="Bintje"  
 /db\_xref="taxon:4113"  
 /clone="cSNAL1K13"  
 /clone\_lib="potato stolon, Cornell University"  
 /tissue\_type="axillary buds of stem explants, swelling  
 stolons"  
 /dev\_stage="1 to 3 days"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2:  
 XhoI; RNA was supplied by Christian Bachem & Beatrix  
 Horvath/Laboratory of Plant Breeding, Dept. of Plant  
 Sciences, Wageningen University, The Netherlands). Total  
 RNA was isolated from developing axillary buds of potato  
 nodal stem cuttings cultured on medium for the  
 introduction of tuber formation as described in Bachem et  
 al. (Plant Journal 1996). Tissue samples were taken of  
 stages corresponding to growing stolons and the early  
 stages of tuber formation."

BASE COUNT 121 a 60 c 74 g 107 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1575 ATTGCTCTCTCTCTCTCTGAT 1597  
 ||||||||||||||||  
 Db 58 ATTGCTCTCTCTCTCTCTGAT 36

RESULT 36  
 BM930440 385 bp mRNA linear EST 13-MAR-2002  
 LOCUS BM930440  
 DEFINITION UI-E-EJ1-ejm-c-11-0-UI-r1 UI-E-EJ1 Homo sapiens cDNA clone  
 UI-E-EJ1-ejm-c-11-0-UI 5', mRNA sequence.  
 ACCESSION BM930440  
 VERSION BM930440.1 GI:19389613  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 385)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565

Email: msouresblue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 REVERSE.

# FEATURES

Source  
 Location/Qualifiers  
 1..385  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-EJ1-ajm-c-11-0-01"  
 /clone\_lib="UI-E-EJ1"  
 /tissue\_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and choroid"  
 /dev\_stage="fetal and adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: pT713-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EJ1 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT713-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAGAGA; lens, CGATTACGGA; eye anterior segment, AATGCCGAT; optic nerve, CCAATPAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."  
 BASE COUNT 156 a 41 c 99 g 89 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 14; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2010 CAAATAGAGAAAAA 2032  
 ||||||||||||||||||  
 Db 356 CAAATAGAGAAAAA 378

RESULT 37  
 BI505360 397 bp mRNA linear EST 08-APR-2002  
 LOCUS BBI70026A10E04.5 Bee Brain Normalized/Subtracted Library, BBI7 Apis  
 DEFINITION mellifera CDNA clone BBI70026A10E04 5', mRNA sequence.  
 ACCESSION BI505360  
 VERSION BI505360.1 GI:15355734  
 KEYWORDS EST.  
 SOURCE honeybee.  
 ORGANISM Apis mellifera  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota;  
 Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
 Apidae; Apis.  
 1 (bases 1 to 397)  
 Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,  
 Pardinas,U., Robertson,H.M., Soares,B. and Robinson,G.E.  
 Annotated expressed sequence tags and cDNA microarrays for studies  
 of brain and behavior in the honey bee  
 Genome Res. 12 (4), 555-566 (2002)  
 JOURNAL 21929762  
 MEDLINE  
 COMMENT Contact: Gene E. Robinson  
 Department of Entomology  
 University of Illinois  
 505 S. Goodwin Ave., Urbana, IL 61801, USA  
 Tel: 217 265 0303

Fax: 217 244 3499  
 Email: generob@life.uiuc.edu  
 This research was funded by the University of Illinois Critical  
 Research Initiatives fund and a Burroughs-Wellcome Trust Innovation  
 Award in Functional Genomics to G.E. Robinson and an NSF  
 Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
 REPEATS IN THE SEQUENCE  
 Low complexity STRAND (+) ELEMENT AT rich LOCATION [220,275];  
 Simple repeat STRAND (+) ELEMENT (TTAA)n LOCATION [281,354];  
 Simple repeat STRAND (+) ELEMENT (A)n LOCATION [357,393].  
 PCR Primers  
 FORWARD: TAAATGACCTCCTATAGG  
 BACKWARD: ATTACCTCCTATAGG  
 Plate: BBI70026A10 row: E column: 04  
 Seq primer: AGCGATACATTTTCACACGGA  
 High quality sequence stop: 397.  
 FEATURES  
 Source  
 Location/Qualifiers  
 1..397  
 /organism="Apis mellifera"  
 /strain="mixed strains of European bees, predominantly  
 A.m. ligustica"  
 /db\_xref="taxon:7460"  
 /clone="BBI70026A10E04"  
 /clone\_lib="Bee Brain Normalized/Subtracted Library, BBI7"  
 /sex="female"  
 /tissue\_type="brain"  
 /dev\_stage="adult worker honey bee"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pT713-Pac; Site\_1: EcoRI;  
 Site\_2: NotI. This BBI7 cDNA library was generated by  
 subtraction of the BBI6 library with 4000 previously  
 sequenced clones. The BBI6 library was contributed by the  
 Soares laboratory and it was constructed and normalized  
 as described by Bonaldo, M.F., Lennon, G. and Soares,  
 M.B. (1996), Genome Research 6(9): 791-806. RNA was  
 prepared from dissected brains of adult worker bees of  
 various ages and various behavioral groups."

BASE COUNT 175 a 47 c 50 g 123 t  
 ORIGIN

Query Match 1.1%; Score 23; DB 13; Length 397;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2010 CAAATAGAGAAAAA 2032  
 ||||||||||||||||||  
 Db 360 CAAATAGAGAAAAA 382

RESULT 38  
 BE109147 408 bp mRNA linear EST 13-JUN-2000  
 LOCUS BE109147/C  
 DEFINITION UI-R-BS1-azb-e-02-0-UI.s1 UI-R-BS1 Rattus norvegicus CDNA clone  
 ACCESSION BE109147  
 VERSION BE109147.1 GI:8501252  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 408)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL 97044477  
 MEDLINE  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250

Fax: 319 335 9565  
Email: mscares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-48,  
>POLY\_A\$Simple\_repeat  
Seq primer: M13 Forward  
POLYA=yes

FEATURES  
Source Location/Qualifiers

1..408  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UT-R-BS1-azd-e-02-0-UT"  
/clone\_1lb="UT-R-BS1"  
/dev\_stage="embryonic 13 dpc"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UT-R-BS1 library is derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu).  
TAG\_L1b=UT-R-BS1  
TAG\_TISSUE=embryo at 13 dpc  
TAG\_SEQ=AATCC"  
BASE COUNT 73 a 102 c 76 g 157 t  
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 408;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2031  
|||||  
Db 23 CCAATATAGAAAAA 1

RESULT 39  
AI095726/c 413 bp mRNA linear EST 05-OCT-1998  
LOCUS g22412.x1 Soares-Pregnant\_Uterus\_NbHPU Homo sapiens cDNA clone  
IMAGE:1697231 3', mRNA sequence.  
ACCESSION AI095726  
VERSION AI095726.1 GI:3434702  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 413)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
Insert Length: 1443 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 405.

FEATURES  
Source Location/Qualifiers  
1..413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1697231"

/clone\_1lb="Soares\_Pregnant\_Uterus\_NbHPU"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15',  
AAGCGAAGAAATCGCGGCCGCTTTTCTTTTCTTTT 3',  
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonafide."  
BASE COUNT 96 a 93 c 105 g 119 t  
ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 413;  
Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2009 CCAATATAGAAAAA 2031  
|||||  
Db 23 CCAATATAGAAAAA 1

RESULT 40  
AM080577/c 419 bp mRNA linear EST 14-OCT-1999  
LOCUS xc33e07.x1 NCI\_CGAP\_Col8 Homo sapiens cDNA clone IMAGE:2586084 3'  
DEFINITION similar to gb:U13315 LACTOYLGLUTATHIONE LYASE (HUMAN);, mRNA sequence.  
ACCESSION AM080577  
VERSION AM080577.1 GI:5035729  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 419)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)  
Seq primer: 40UP from Gibco.

FEATURES  
Source Location/Qualifiers

1..419  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2586084"  
/clone\_1lb="NCI\_CGAP\_Col8"  
/tissue\_type="moderately differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.26 kb. Library constructed by Life Technologies. Normalized versions of this library named NCI-CGAP\_Col9 (Cot 50), NCI\_CGAP\_Col20 (Cot 500), and NCI-CGAP\_Col21 (Cot >500)."  
BASE COUNT 130 a 72 c 69 g 148 t  
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032  
 |||||||  
 Db 31 CAATATAGAAAAA 9

RESULT 41  
 BE107411/c 420 bp mRNA linear EST 13-JUN-2000  
 LOCUS UI-R-BS1-ayt-d-07-0-UI-s1 UI-R-BS1 Rattus norvegicus cDNA clone  
 DEFINITION UI-R-BS1-ayt-d-07-0-UI 3', mRNA sequence.  
 ACCESSION BE107411 GI:8495516  
 VERSION BE107411.1  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 420)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a donatide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 normalized embryo at 13 dpc library cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com) The following repetitive  
 elements were found in this cDNA sequence: 1-48,  
 >POLY\_ASimple-repeat  
 Seq primer: M13 Forward  
 POLYA=yes.

FEATURES  
 source  
 1. 420  
 Location/Qualifiers  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-BS1-ayt-d-07-0-UI"  
 /clone\_lib="UI-R-BS1"  
 /dev\_stage="embryonic 13 dpc"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pTZ19-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-BS1  
 library is derived from 13 dpc whole embryo tissue. For a  
 detailed description of the library from which this clone  
 was derived, please visit our web site at  
 ratest.eng.uiowa.edu.  
 TAG\_LIB=UI-R-BS1  
 TAG\_TISSUE=embryo at 13 dpc  
 TAG\_SEQ=ATATC"  
 BASE COUNT 76 a 104 c 80 g 160 t  
 ORIGIN  
 Query Match 1.1%; Score 23; DB 10; Length 420;  
 Best Local Similarity 100.0%; Pred. NO. 1.7e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2031  
 |||||||  
 Db 23 CCAATATAGAAAAA 1

RESULT 42  
 A1195993/c 423 bp mRNA linear EST 14-OCT-1998  
 LOCUS u63b09 x1 Sugano mouse liver tlia Mus musculus cDNA clone  
 DEFINITION IMAGE:1887065 3' similar to gb:M16961 ALPHA-2-HS-GLYCOPROTEIN  
 PRECURSOR (HUMAN);, mRNA sequence.  
 ACCESSION A1195993  
 VERSION A1195993  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 423)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M.,  
 Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,  
 Thelsting,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:971389  
 Seq primer: custom primer used  
 High quality sequence stop: 409.

FEATURES  
 source  
 1. 423  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1887065"  
 /clone\_lib="Sugano mouse liver tlia"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: liver; Vector: pME18S-FL3; Site\_1: DraIII  
 (CACTGTTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
 was primed with an oligo(dT) primer  
 [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was  
 ligated to a DraIII adaptor (TTGTGGCTCTG), digested  
 and cloned into distinct DraIII sites of the pME18S-FL3  
 vector (5' site CACTGTTG, 3' site CACCATGTG). XhoI should  
 be used to isolate the cDNA insert. Size selection was  
 performed to exclude fragments <1.5kb. Library  
 constructed by Dr. Sumio Sugano (University of Tokyo  
 Institute of Medical Science). Custom primers for  
 sequencing: 5' end primer CTGTGCTTAAAGCTGG and 3' end  
 primer CGACCTGCACCTCAGACAC."

BASE COUNT 89 a 103 c 106 g 125 t  
 ORIGIN  
 Query Match 1.1%; Score 23; DB 9; Length 423;  
 Best Local Similarity 100.0%; Pred. NO. 1.7e+03;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032  
 |||||||  
 Db 23 CAATATAGAAAAA 1

RESULT 43  
 W32300/c 437 bp mRNA linear EST 10-OCT-1996  
 LOCUS

DEFINITION zc05g05.s1 Soares-Parathyroid-tumor-NBHPA Homo sapiens cDNA clone IMAGE:321464 3' similar to gb:D13315 LACTOYLGLUTATHIONE LYASE (HUMAN); mRNA sequence.

ACCESSION W33300

VERSION W33300.1 GI:1313311

KEYWORDS EST.

SOURCE human

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 437)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rikhi, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The Washu-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1429 Std Error: 0.00  
Seq primer: mob, REGA+ET  
High quality sequence stop: 374.

FEATURES

Source

Location/Qualifiers

1..437

/organism="Homo sapiens"

/db\_xref="GDB:1259122"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:321464"

/clone\_id="Soares-Parathyroid-tumor-NBHPA"

/tissue\_type="parathyroid tumor"

/dev\_stage="adult"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: parathyroid gland; Vector: pTZ19 (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer

15'-TGTTACCAATCTGACGTGGAGCGCGACACCAATTTTTTTTTTTTTTTT  
TTTT-3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldi. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 137 a 75 c 73 g 151 t 1 others

ORIGIN

Query Match 1.1%; Score 23; DB 14; Length 437;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2010 CAATATAGAAAAA 2032  
|||||

DB 25 CAATATAGAAAAA 3

RESULT 44  
A1952475/c 443 bp mRNA linear EST 09-MAR-2000  
LOCUS w74c04.x1 NCI CGAP Ov38 Homo sapiens cDNA clone IMAGE:2549382 3'  
DEFINITION similar to gb:D13315 LACTOYLGLUTATHIONE LYASE (HUMAN); mRNA  
sequence.  
ACCESSION A1952475.  
VERSION A1952475.1 GI:5744785  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 443)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Jeff Struwing, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA library preparation: Life Technologies, Inc. cDNA library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert Length: 2122 Std Error: 0.00  
Seq primer: -40up from G1bco  
High quality sequence stop: 396.

FEATURES

Source

Location/Qualifiers

1..443

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:2549382"

/clone\_id="NCI-CGAP-Ov38"

/tissue\_type="normal epithelium"

/lab\_host="DH10B"

/note="Organ: ovary; Vector: PCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

BASE COUNT 139 a 77 c 75 g 152 t

ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 443;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2010 CAATATAGAAAAA 2032  
|||||

DB 27 CAATATAGAAAAA 5

RESULT 45  
B0013210/c 446 bp mRNA linear EST 26-MAR-2002  
LOCUS UT-1-BC1P-ayt-d-07-0-01.s1 NCI-CGAP\_P13 Homo sapiens cDNA clone  
DEFINITION UT-1-BC1P-ayt-d-07-0-01.s1, mRNA sequence.

ACCESSION B0013210

VERSION B0013210.1 GI:19738111

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 446)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Dr. Steven Brown  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu  
Seq primer: M13 FORWARD  
POLYA-yes.

FEATURES

Source

Location/Qualifiers

1..446

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="U1-1-BC1p-ayt-d-07-0-UI"
/clone_lib="NCI-CGAP_P13"
/tissue_type="placenta"
/dev_stage="8-9 weeks"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Placenta; Vector: pT73-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoR I; Site_2: Not I;
NCI-CGAP_P13 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGGA, for
additional information, contact: Bento Soares,
bento_soares@uiowa.edu
TAG_LIB=U1-1-BC1p
TAG_TISSUE=placenta human 8 week
TAG_SEQ=GA"

BASE COUNT      94 a      105 g      128 t
ORIGIN

Query Match      1.18; Score 23; DB 14; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032
|||||
Db 23 CAATAGAGAAAAA 1

RESULT 46
BQ427521      465 bp      mRNA      linear      EST 24-MAY-2002
LOCUS
DEFINITION
AGENCOURT_7919274 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6160346
5', mRNA sequence.
ACCESSION
BQ427521
VERSION
BQ427521.1 GI:21166597
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 465)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@femail.nih.gov
Tissue procurement: ATCC/DCMD/PTP
CDNA library preparation: Life Technologies, Inc.
CDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13510 row: 0 column: 03
High quality sequence stop: 464.
Location/Qualifiers
1. 465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6160346"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

```

```

Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT      155 a      83 c      82 g      144 t      1 others
ORIGIN

Query Match      1.1%; Score 23; DB 14; Length 465;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032
|||||
Db 442 CAATAGAGAAAAA 464

RESULT 47
BQ194968      481 bp      mRNA      linear      EST 30-APR-2002
LOCUS
DEFINITION
U1-R-CNI-cmd-k-19-0-UI.s1 U1-R-CNI Rattus norvegicus cDNA clone
U1-R-CNI-cmd-k-19-0-UI 3', mRNA sequence.
ACCESSION
BQ194968
VERSION
BQ194968.1 GI:20370519
KEYWORDS
EST.
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 481)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
97044477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscares@iue.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonalde poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized fundus library cDNA library preparation: M.B. Soares Lab
Genetics (www.reagen.com)
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1. 481
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="U1-R-CNI-cmd-k-19-0-UI"
/clone_lib="U1-R-CNI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The U1-R-CNI
library is a subtracted library derived from the following
pool of seven normalized rat libraries: normalized rat
bladder vesicles, normalized rat penis, normalized rat
bladder, normalized rat cervix, normalized rat brown
adipose, normalized rat fundus, and normalized rat
salivary gland. It was constructed according to the
procedure described by Bonaldo, Lennon & Soares (genome
Research volume 6: 791-806, 1996). For construction of
the CNI library, plasmid DNA from the pool of normalized
libraries was electroporated into competent bacteria for
the production of single-stranded circular DNA. This was
then used as a tracer in a subtractive hybridization with

```

a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Urogen Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and NN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries C40 and C41 corresponding to plates R-C40-AMV through R-C40-AS, R-C40-AZX through R-C40-BAZ, R-C40-BEE through R-C40-BHY, R-C40-BJS, R-C40-BKE, R-C40-BKG-H, R-C40-BKJ-K, R-C40-BKE through R-C40-BKS, R-C40-BKV-V, R-C40-BLY through R-C40-BMA, R-C40-BMC through R-C40-BME, R-C40-BNS, R-C40-BOB through R-C40-BOJ, R-C40-BPA through R-C40-BPG, R-C41-BBA through R-C41-BDA, R-C41-BHZ through R-C41-BIF, R-C41-BIR, R-C41-BJT through R-C41-BKB, R-C41-BKD, R-C41-BK, R-C41-BKI, R-C41-BKT, R-C41-BLF, R-C41-BLH through R-C41-BLN, R-C41-BLS, R-C41-BLV-V, R-C41-BNR, and R-C41-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, CW0s, CX0s and normalized libraries CS0, CT0, CU0, CW0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAV, R-CU0s-CBP through R-CU0s-CCB, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN, R-CU0-BDQ through R-CU0-BYL, R-CW0-BVQ through R-CW0-BMP, R-CW0-BXN through R-CW0-BXO, R-CX0-BWQ through R-CX0-BXN. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CVO and 4,000 from normalized eye library CV1) corresponding to plates R-CVO-BSE through R-CVO-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BVO and BV0P (7.9.5 kb CDNA library fraction from rat whole embryo), and BX0 (0.5-7kb CDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0P-AOI through R-BV0P-ACX, and R-BX0-AOX through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CY0, CZ0, DA0, DB0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYL, R-DA0-BZB-C, R-DD0-BVJ through R-DD0-BZD, R-DD0-BZD through R-DD0-BZQ, R-DD0-BZQ through R-DD0-BZA, R-DD0-BZI through R-DD0-CAI, R-DD0-CBA, R-DD0-BZR through R-DD0-CAI. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CNO, that makes up the tracer. The corresponding plates are R-CNO-BKV through R-CNO-BLD, R-CNO-BIG, R-CNO-BIP through R-CNO-BIR, R-CNO-BIR, R-CNO-BLM-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CNO pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkx-c-06-0-UI, bkx-c-08-0-UI, bkx-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CNO clone (corresponding to the address bkx-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

FEATURES

source

1.484

Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-BJ0P-afd-c-06-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-BJ0P library is a subtracted library derived from the UI-R-AA1, UI-R-RAB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research

2010 CAATATAGCAAAAAAAAAAAAAA 2032

DB 23 CAATATAGCAAAAAAAAAAAAAA 1

RESULT 48

AM434101/c

LOCUS

DEFINITION

UI-R-BJ0P-afd-c-06-0-UI s1 UI-R-BJ0P Rattus norvegicus cDNA clone

ACCESSION

AM434101

VERSION

AM434101.1 GI:6965408

KEYWORDS

EST.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 484)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

COMMENT

CONTACT: Soares, MB

9704477

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msocres@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized AV canal at 16.5 dpc library cDNA library preparation:

M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-48,

>POLY\_A#simple-repeat

Seq primer: M13 Forward

POLYA=yes.

Query Match 1.1%; Score 23; DB 14; Length 481;

Best Local Similarity 100.0%; Pred. No. 1.6e+03;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

117 a 103 c 124 g 137 t

TAG-TIB-UI-R-CNI

TAG-TISSUE-fundus

TAG\_SEQ=TTCCG"

6: 791-806, 1996.  
TAG\_LIB=01-R-BJOP  
TAG\_TISSUE=AV canal at 16.5 dpc  
TAG\_SEQ=GACC"

BASE COUNT

86 a 113 c 100 g 185 t

ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATAGAGAAAAA 2031

Db 23 CCAATAGAGAAAAA 1

RESULT 49  
BF062672/ c 484 bp mRNA linear EST 16-OCT-2000  
LOCUS 7653602.x1 NCI\_CGAP\_CO16 Homo sapiens cDNA clone IMAGE:3320650 3',  
DEFINITION mRNA sequence.  
ACCESSION BF062672  
VERSION BF062672.1 GI:10821582  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 484)  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -400p from Gibco  
High quality sequence stop: 377.  
Location/Qualifiers

JOURNAL COMMENT

FEATURES

source

1. 484  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP\_CO16"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_CO10 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1057416-1061255, and 1144584-1145351)  
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT

149 a 87 c 80 g 168 t

ORIGIN

Query Match 1.1%; Score 23; DB 12; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CCAATAGAGAAAAA 2032

Db 26 CCAATAGAGAAAAA 4

RESULT 50  
AA809972/ c 497 bp mRNA linear EST 19-FEB-1998  
LOCUS 0a95f12.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1320047 3',  
DEFINITION similar to gb:U13315 LACTOYLGLUTATHIONE LYASE (HUMAN);, mRNA  
sequence.  
ACCESSION AA809972  
VERSION AA809972.1 GI:2879378  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 497)  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert length: 2392 Std Error: 0.00  
Seq primer: -40m13 fwd. Er from Amersham  
High quality sequence stop: 370.  
Location/Qualifiers

JOURNAL COMMENT

FEATURES

source

1. 497  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAGTGGAGCGCCGCTCTTTTCTTTTCTTTT-3'  
]. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73D vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT

154 a 93 c 84 g 166 t

ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 497;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CCAATAGAGAAAAA 2032

Db 23 CCAATAGAGAAAAA 1

Search completed: January 11, 2003, 01:21:55  
Job time : 3172 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:57:46 ; Search time 80 Seconds  
(without alignments)  
7789.590 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032 1 ggcctcgagccagagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO\_NDC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 segs, 15338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 14

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2032	100.0	2032	4	US-09-045-284A-1
2	2032	100.0	2032	4	US-09-190-911-2
3	37	1.8	37	4	US-09-045-284A-7
4	37	1.8	37	4	US-09-190-911-7
5	22	1.1	3138	1	US-07-867-106-4
6	21	1.0	1493	6	5340934-5
7	20	1.0	295	1	US-08-171-385-16
8	20	1.0	295	3	US-08-361-441B-16
9	20	1.0	362	1	US-08-171-385-12
10	20	1.0	362	3	US-08-361-441B-12
11	20	1.0	387	4	US-09-247-155-152
12	20	1.0	1869	4	US-09-336-536-26
13	20	1.0	2429	4	US-09-360-545-68
14	20	1.0	55827	4	US-09-813-133A-3

ALIGNMENTS

RESULT 1  
US-09-045-284A-1  
Sequence 1 Application US/09045284A  
Patent No. 6285192  
GENERAL INFORMATION:  
APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107US1  
CURRENT APPLICATION NUMBER: US/09/045,284A  
CURRENT FILING DATE: 1998-03-20  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FASTSQ For windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2032  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-045-284A-1

Query Match 100.0%; Score 2032; DB 4; Length 2032;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGCTCGAGGCGCAGGATGCTCCAGTCGCGGGGAAATGCTTCATTTGCTTCCGAG	60
DB	1	GGCTCGAGGCGCAGGATGCTCCAGTCGCGGGGAAATGCTTCATTTGCTTCCGAG	60
QY	61	CCGACCTCAAGCAGTCTCCACCCCTTGAGTCTCAGAGTGTAAAGCTGTACTTCA	120
DB	61	CCGACCTCAAGCAGTCTCCACCCCTTGAGTCTCAGAGTGTAAAGCTGTACTTCA	120
QY	121	CAGCTTCCTGGGAGCGAGTGTCTTCAAGCCGCTTTCAGAGTCTTCCAGTTCAGC	180
DB	121	CAGCTTCCTGGGAGCGAGTGTCTTCAAGCCGCTTTCAGAGTCTTCCAGTTCAGC	180
QY	181	AATGCTACTGCTTAAATAATGAGCTGCTGTTTGGTTTCCAGATGGCATCTT	240
DB	181	AATGCTACTGCTTAAATAATGAGCTGCTGTTTGGTTTCCAGATGGCATCTT	240
QY	241	GGCTTATTTTCCATGATGACAGCCACACATCACTCCCTGTCTATGAAAGACAGCC	300
DB	241	GGCTTATTTTCCATGATGACAGCCACACATCACTCCCTGTCTATGAAAGACAGCC	300
QY	301	CGAGCGATGACAGTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	360
DB	301	CGAGCGATGACAGTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	360
QY	361	GCTTTTGGGAGCAGCCAGATGTTTCTTCACTGATGAGCCGCTGACAGTGTGAT	420
DB	361	GCTTTTGGGAGCAGCCAGATGTTTCTTCACTGATGAGCCGCTGACAGTGTGAT	420
QY	421	GACCTTCAGCAGACAGCCGCTGATGCTGACATGAGTGTGCGGATCTATAGGCGC	480
DB	421	GACCTTCAGCAGACAGCCGCTGATGCTGACATGAGTGTGCGGATCTATAGGCGC	480
QY	481	CGCTTCCTGTCGATGATGAGCTCTTTGATGCTTCAATGAACTGTCCTGCGAGACA	540
DB	481	CGCTTCCTGTCGATGATGAGCTCTTTGATGCTTCAATGAACTGTCCTGCGAGACA	540
QY	541	GTCACGCTCTTCACTGAGAGACAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTG	600
DB	541	GTCACGCTCTTCACTGAGAGACAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTG	600
QY	601	CATCCCAAGATGAAATCATCCCGGGGTCACAGAGTCTGAGAGCAACAGCC	660
DB	601	CATCCCAAGATGAAATCATCCCGGGGTCACAGAGTCTGAGAGCAACAGCC	660
QY	661	CTTTGAGTGTGAGAGAGCGCTGCTCTTCAAGCAGCTGAGTGTCTCAAGAGGTGCG	720
DB	661	CTTTGAGTGTGAGAGAGCGCTGCTCTTCAAGCAGCTGAGTGTCTCAAGAGGTGCG	720
QY	721	CTTCTTCAAGCTGAGTGTCTTCAAGCAGCTGAGTGTCTTCAAGCAGCTGAGTGT	780
DB	721	CTTCTTCAAGCTGAGTGTCTTCAAGCAGCTGAGTGTCTTCAAGCAGCTGAGTGT	780
QY	781	CGTCAAGCTGTGTCGAGAGCCCGGGCGTGTTCGCTTCCGAGAACGCAAGAGGAGA	840
DB	781	CGTCAAGCTGTGTCGAGAGCCCGGGCGTGTTCGCTTCCGAGAACGCAAGAGGAGA	840
QY	841	TCTCATGATTGACAGTCGATTTGTGATGGGCGAGCATGAGCAAACTCAAGAGAGAGA	900

Db 841 TCTCATATTTGACAGTCGCTGATGGATGGGAGCATATGACCAAACTCAGAGAGAGG 900  
QY 901 CCAACCTTACTATGTATGAGGTTCATCTGCGCAAGCCAGCTGAGATCTACAAACCAT 960  
Db 901 CCAACCTTACTATGTATGAGGTTCATCTGCGCAAGCCAGCTGAGATCTACAAACCAT 960  
QY 961 CCAATCTCTTCCCAAGGCGCTGAGAGACCTACCTGCTTGTGCTGTATGAGAGACCTGGC 1020  
Db 961 CCAATCTCTTCCCAAGGCGCTGAGAGACCTACCTGCTTGTGCTGTATGAGAGACCTGGC 1020  
QY 1021 TCGAGCCCTGTGGCCAGACTTCCGAATGTATGATTTGGATTTGGATTTCTTGGC 1080  
Db 1021 TCGAGCCCTGTGGCCAGACTTCCGAATGTATGATTTGGATTTGGATTTCTTGGC 1080  
QY 1081 CCAATCTTCAAGCTGGGTGATACATCACCCAGGCAAGGCGATGGGTGACCACTTT 1140  
Db 1081 CCAATCTTCAAGCTGGGTGATACATCACCCAGGCAAGGCGATGGGTGACCACTTT 1140  
QY 1141 CCACACAATGCCAGGATCCCTTATGTCTCCAGGCTTGGCGCTGCTTGGCCTTA 1200  
Db 1141 CCACACAATGCCAGGATCCCTTATGTCTCCAGGCTTGGCGCTGCTTGGCCTTA 1200  
QY 1201 TGAAGAAGTTTCTGAGCTTGAGAAAGCCTGTGGCATGCCATGAATTTGCTGGCTACCG 1260  
Db 1201 TGAAGAAGTTTCTGAGCTTGAGAAAGCCTGTGGCATGCCATGAATTTGCTGGCTACCG 1260  
QY 1261 CCAAGTCAGATTCGAGAAAGAAAGAAACCTGTGCTGATCTCTGCTACCTGGAC 1320  
Db 1261 CCAAGTCAGATTCGAGAAAGAAAGAAACCTGTGCTGATCTCTGCTACCTGGAC 1320  
QY 1321 TGTCCCTGAGCAATCCACTAAGAGGGTTGAGAGGCTTGTGCTGCCACCTGATGACGC 1380  
Db 1321 TGTCCCTGAGCAATCCACTAAGAGGGTTGAGAGGCTTGTGCTGCCACCTGATGACGC 1380  
QY 1381 TCAATCTCTTCTGATGATCTTCTGAGCCTTGCCTACATCTCTGAGCCTTAACCTAC 1440  
Db 1381 TCAATCTCTTCTGATGATCTTCTGAGCCTTGCCTACATCTCTGAGCCTTAACCTAC 1440  
QY 1441 TCTGTGGGATTCACACTGATGATGATGTGTCCACAGCTGCTCAAGCAGAGAGACTTTT 1500  
Db 1441 TCTGTGGGATTCACACTGATGATGATGTGTGTCCACAGCTGCTCAAGCAGAGAGACTTTT 1500  
QY 1501 GTGTCCATGCTTGTGTATGAGAAACAGACTGGGAACTTATGTGAGCAGACATATCCAC 1560  
Db 1501 GTGTCCATGCTTGTGTATGAGAAACAGACTGGGAACTTATGTGAGCAGACATATCCAC 1560  
QY 1561 CAGTGAACAGAGGATGCTCTTCTTCTTGTGATCTTCTGCTGGGACAGACTTCA 1620  
Db 1561 CAGTGAACAGAGGATGCTCTTCTTCTTGTGATCTTCTGCTGGGACAGACTTCA 1620  
QY 1621 AGACTTTGTGGCTGAGGCTTATTAAGCAGCAGACAGTATCAGTGAATTTGATCCATAA 1680  
Db 1621 AGACTTTGTGGCTGAGGCTTATTAAGCAGCAGACAGTATCAGTGAATTTGATCCATAA 1680  
QY 1681 ACCGCCCTGTCCATGCTGCCAATGGGGAATGSAATCTTCCAAAAGGCTCACAG 1740  
Db 1681 ACCGCCCTGTCCATGCTGCCAATGGGGAATGSAATCTTCCAAAAGGCTCACAG 1740  
QY 1741 ATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGATTCACGATGATTAAGAGAA 1800  
Db 1741 ATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGATTCACGATGATTAAGAGAA 1800  
QY 1801 GTGGGAACAGAGTTGATGCTTACTTATGAGCTTGACATACAGCTATCGTAATCA 1860  
Db 1801 GTGGGAACAGAGTTGATGCTTACTTATGAGCTTGACATACAGCTATCGTAATCA 1860  
QY 1861 AATGTGAACAAATCTCTGACAAAGAGCAGCTTATAGTTCAAGGCTCCGGGC 1920  
Db 1861 AATGTGAACAAATCTCTCTGACAAAGAGCAGCTTATAGTTCAAGGCTCCGGGC 1920  
QY 1921 TGCATTTGAATATCATCTCCCTCTGCAATTTCCATCACATGAAGACTTTGACCTGTG 1980  
Db 1921 TGCATTTGAATATCATCTCCCTCTGCAATTTCCATCACATGAAGACTTTGACCTGTG 1980

Db 1921 TGCATTTGAATATCATCTCCCTCTGCAATTTCCATCACATGAAGACTTTGACCTGTG 1980  
QY 1981 AAGTGGCATCTGTTAATACATAAATTTCCAAATATAGAAAAA 2032  
Db 1981 AAGTGGCATCTGTTAATACATAAATTTCCAAATATAGAAAAA 2032  
RESULT 2  
US-09-190-911-2  
; Sequence 2, Application US/09190911  
; Patent No. 6365365  
; GENERAL INFORMATION:  
; APPLICANT: Bistrup, Annette  
; APPLICANT: Rosen, Steven D.  
; APPLICANT: Tangemann, Kirsten  
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
; FILE REFERENCE: 6510-107CIP  
; CURRENT APPLICATION NUMBER: US/09/190, 911  
; EARLIER FILING DATE: 1998-11-12  
; EARLIER APPLICATION NUMBER: 09/045,284  
; EARLIER FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 2032  
; TYPE: DNA  
; ORGANISM: H. sapiens  
US-09-190-911-2  
Query Match 100.0%; Score 2032; DB 4; Length 2032;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTGGAGGCGCGATGCTCCAGCTGAGGGAATAAGCTTCATTTGCTTCCAG 60  
Db 1 GGTGGAGGCGCGATGCTCCAGCTGAGGGAATAAGCTTCATTTGCTTCCAG 60  
QY 61 CCCACCTAAGCAGTCTCCCAACCCCTTGAATCTCAGAGGTAAAGCTTAATTCA 120  
Db 61 CCCACCTAAGCAGTCTCCCAACCCCTTGAATCTCAGAGGTAAAGCTTAATTCA 120  
QY 121 CAGCTTCCGAGGAGGAGTCTTCTCAAGCCGCTGCAAGGCTTCCACTCAGAC 180  
Db 121 CAGCTTCCGAGGAGGAGTCTTCTCAAGCCGCTGCAAGGCTTCCACTCAGAC 180  
QY 181 AATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTTCTGATTTCCAGATGGCATCTT 240  
Db 181 AATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTTCTGATTTCCAGATGGCATCTT 240  
QY 241 GCTCTATTTCTCCAGATGATGACCAACATGACGCTCCCTGTCTATGAGAGGACAGC 300  
Db 241 GCTCTATTTCTCCAGATGATGACCAACATGACGCTCCCTGTCTATGAGAGGACAGC 300  
QY 301 CGAGCGCATGACGCTGCTGCTGCTTCTGCTGAGCGCTGCTGCTTCTTGTGGGCA 360  
Db 301 CGAGCGCATGACGCTGCTGCTGCTTCTGCTGAGCGCTGCTGCTTCTTGTGGGCA 360  
QY 361 GCTTTTGGGCGACACCCAGATGTTTCTTACTGATGAGCCCGCTGGCAGCTGTGAT 420  
Db 361 GCTTTTGGGCGACACCCAGATGTTTCTTACTGATGAGCCCGCTGGCAGCTGTGAT 420  
QY 421 GACCTTCAAGCAGACCGCTGGATGCTGACATGCTGTGGGGATTCGATACGGGC 480  
Db 421 GACCTTCAAGCAGACCGCTGGATGCTGACATGCTGTGGGGATTCGATACGGGC 480  
QY 481 GCTCTTCTGTCGACATGAGGCTTTTGANGCATAGAACTGTCCCGAGACA 540  
Db 481 GCTCTTCTGTCGACATGAGGCTTTTGANGCATAGAACTGTCCCGAGACA 540  
QY 541 GTCCAGCCTCTTTCAGTGGGAGAAAGCCGGGCTGTGTTCTGTGACCTGCTGTGACAT 600  
Db 541 GTCCAGCCTCTTTCAGTGGGAGAAAGCCGGGCTGTGTTCTGTGACCTGCTGTGACAT 600

```
QY 601 CATCCCAAGATGAAATCATATCCCGGGGCTCAGTGCAGAGCTCTCTGTGCACTCAACAGCC 660
Db 601 CATCCCAAGATGAAATCATATCCCGGGGCTCAGTGCAGAGCTCTCTGTGCACTCAACAGCC 660
QY 661 CTTTGAGGTGTGAGAGAGGCTCGCGCTCCAGAGCCAGCTGTGTGCTCAAGAGAGTGGC 720
Db 661 CTTTGAGGTGTGAGAGAGGCTCGCGCTCCAGAGCCAGCTGTGTGCTCAAGAGAGTGGC 720
QY 721 CTCTTCAACCTGCAAGTCTCTTACCCGCTGCTGAAGAGCCCTCCCTCAACTGCAATAT 780
Db 721 CTCTTCAACCTGCAAGTCTCTTACCCGCTGCTGAAGAGCCCTCCCTCAACTGCAATAT 780
QY 781 CGTGCACCTGTGCGGGAGCCCGGGGCGCTGTCCGTTCCCGAGAGACGACAAAGGGAGGA 840
Db 781 CGTGCACCTGTGCGGGAGCCCGGGGCGCTGTCCGTTCCCGAGAGACGACAAAGGGAGGA 840
QY 841 TGTCAATGATGACATGCGCATTTGTATGAGGAGCAGATGAGCAGAACTCAAGAGAGGA 900
Db 841 TGTCAATGATGACATGCGCATTTGTATGAGGAGCAGATGAGCAGAACTCAAGAGAGGA 900
QY 901 CCAACCCCTACATGTGATGAGCAGGTGATCTGCCAAGCCAGCTGAGATCTCAAGACCAT 960
Db 901 CCAACCCCTACATGTGATGAGCAGGTGATCTGCCAAGCCAGCTGAGATCTCAAGACCAT 960
QY 961 CCAGTCTTGGCCCAAGGCGCTGACAGAAAGCTACCTGCTGTGCGCTATGAGAGCTGGC 1020
Db 961 CCAGTCTTGGCCCAAGGCGCTGACAGAAAGCTACCTGCTGTGCGCTATGAGAGCTGGC 1020
QY 1021 TCGAGCCCTGTGCGCCAGACTTCCCGAATGATGATTCGTGGAGATGAAATCTTGGC 1080
Db 1021 TCGAGCCCTGTGCGCCAGACTTCCCGAATGATGATTCGTGGAGATGAAATCTTGGC 1080
QY 1081 CCATCTTACAGCCTGGGTGATACATACATACCCAGAGGAGGAGGATGGGTGACCAAGCTT 1140
Db 1081 CCATCTTACAGCCTGGGTGATACATACATACCCAGAGGAGGAGGATGGGTGACCAAGCTT 1140
QY 1141 CCACAAATGCGAGGAGTGGCCCTTAATGTCTCCAGGCTTGGGCTTGGCTTGGCCTTA 1200
Db 1141 CCACAAATGCGAGGAGTGGCCCTTAATGTCTCCAGGCTTGGGCTTGGCTTGGCCTTA 1200
QY 1201 TGAAGAGTTCCTGACCTTCAAGAAAGCCTGTGCGATGCCATGATATTCGAGGCTACCG 1260
Db 1201 TGAAGAGTTCCTGACCTTCAAGAAAGCCTGTGCGATGCCATGATATTCGAGGCTACCG 1260
QY 1261 CCACGTCAGATCTGACAGAGAGAGAAACCTGTCTGATGATCTTCTGTACTCTGAGC 1320
Db 1261 CCACGTCAGATCTGACAGAGAGAGAAACCTGTCTGATGATCTTCTGTACTCTGAGC 1320
QY 1321 TGTCCCTAGCAATCCACTAAGAGGTTGAGAGGCTTCTCTCCACCTGTGTCAAGC 1380
Db 1321 TGTCCCTAGCAATCCACTAAGAGGTTGAGAGGCTTCTCTCCACCTGTGTGTCAAGC 1380
QY 1381 TCAAGTCACTTCTCTGATGATGCTTTCGAGCCTTGCCTACATCTGAGGCTTAACATCAATG 1440
Db 1381 TCAAGTCACTTCTCTGATGATGCTTTCGAGCCTTGCCTACATCTGAGGCTTAACATCAATG 1440
QY 1441 TCTGTGGGTATCAGACTGAGTGTGATGTGTCCACAGCTGCTCAAGCAGAGAGACTTTT 1500
Db 1441 TCTGTGGGTATCAGACTGAGTGTGATGTGTCCACAGCTGCTCAAGCAGAGAGACTTTT 1500
QY 1501 GTGTCCATGCTGTGTGTAGAGAAAGAGCTGGGAGCTTGTGTGAGAGAGACATCCAC 1560
Db 1501 GTGTCCATGCTGTGTGTAGAGAAAGAGCTGGGAGCTTGTGTGAGAGAGACATCCAC 1560
QY 1561 CAGTGAAGAGAGGATGATCTCTTCTTCTTCTGATCTTCTCTCTGAGGAGACTTCAAG 1620
Db 1561 CAGTGAAGAGAGGATGATCTCTTCTTCTTCTGATCTTCTCTCTGAGGAGACTTCAAG 1620
QY 1621 AGACTTGTGGCTGGAGGCTATTAAGCAGACAGTATCAAGTGAATGATCAATAA 1680
Db 1621 AGACTTGTGGCTGGAGGCTATTAAGCAGACAGTATCAAGTGAATGATCAATAA 1680

QY 1681 ACCTCCCTGTCCACATCTTCCCATGGGAGTGAATGATCTTTCACCAAGAGACTCACACG 1740
Db 1681 ACCTCCCTGTCCACATCTTCCCATGGGAGTGAATGATCTTTCACCAAGAGACTCACACG 1740
QY 1741 ATTTTCACAGAGATGCAATTTCTGAGCCCTTGAGTTCCTCAAGTGTCAAGAGAGAA 1800
Db 1741 ATTTTCACAGAGATGCAATTTCTGAGCCCTTGAGTTCCTCAAGTGTCAAGAGAGAA 1800
QY 1801 GTGGGAACAAGTGTGATGCTTCTTATGAGCTTGACCATCAGCTATTCGTAATCAGA 1860
Db 1801 GTGGGAACAAGTGTGATGCTTCTTATGAGCTTGACCATCAGCTATTCGTAATCAGA 1860
QY 1861 AATATGAACAAATCTCTGACAAAGACAGAGCTTATGATGACAGGAGTCCCTGGG 1920
Db 1861 AATATGAACAAATCTCTGACAAAGACAGAGCTTATGATGACAGGAGTCCCTGGG 1920
QY 1921 TGAATTTGAATATCACTTCCCTCTGCAATTTCCATCAGATGAGAGACTTTGACCTGTG 1980
Db 1921 TGAATTTGAATATCACTTCCCTCTGCAATTTCCATCAGATGAGAGACTTTGACCTGTG 1980
QY 1981 AAGTGCATCTGTATTAATAATTCGCAATTAAGAAAAA 2032
Db 1981 AAGTGCATCTGTATTAATAATTCGCAATTAAGAAAAA 2032

RESULT 3
US-09-045-284A-7
; Sequence 7, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-7

Query Match 1.8%; Score 37; DB 4; Length 37;
Best Local Similarity 100.0%; Pred No. 4.3e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 AACTCAAGAGAGAGGACCAACCTACTATGATGATGC 920
Db 1 AACTCAAGAGAGAGGACCAACCTACTATGATGATGC 37

RESULT 4
US-09-190-911-7
; Sequence 7, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
```

TYPE: DNA  
ORGANISM: H. sapiens  
US-09-190-911-7

Query Match  
Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 AACTCAAGAGAGAGACCACTACTATGTGATGC 920  
|||||  
DB 1 AACTCAAGAGAGAGACCACTACTATGTGATGC 37

RESULT 5  
US-07-867-106-4/c  
Sequence 4, Application US/07867106  
Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B  
APPLICANT: Chang, Andy C M  
APPLICANT: Williams, Keith L  
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular  
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris  
STREET: One Liberty Place 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187  
APPLICATION NUMBER: PCT/A090/00530  
FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Feehey, Joanne Longo  
REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-9100  
TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3138 base pairs

TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO  
US-07-867-106-4

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATATAGAAAAA 2032  
|||||  
DB 2030 AATATAGAAAAA 2009

RESULT 6  
5340934-5

Patent No. 5340934  
APPLICANT: TERMININE, JOHN D., YOUNG, MARIAN F., FISHER, LARRY W.

ROBEY, PAMELA G.  
TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS  
NUMBER OF SEQUENCES: 13  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/432,044  
FILING DATE: 03-NOV-1989  
SEQ ID NO: 5  
LENGTH: 1493

Query Match  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAAAAA 2032  
|||||  
DB 1440 AATAGAAAAA 1460

RESULT 7  
US-08-171-385-16/c  
Sequence 16, Application US/08171385  
Patent No. 5527884

GENERAL INFORMATION:

APPLICANT: Mary E. Russell  
APPLICANT: Ulrike Utans  
TITLE OF INVENTION: Mediators of Chronic Allotraft  
TITLE OF INVENTION: Rejection  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,385

FILING DATE:  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 05433/006001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:  
LENGTH: 295

TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-171-385-16

Query Match  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 AATAGAAAAA 2032  
|||||  
DB 27 AATAGAAAAA 8

RESULT 8  
US-08-361-441B-16/c  
Sequence 16, Application US/08361441B  
Patent No. 6077948  
GENERAL INFORMATION:  
APPLICANT: Russell, Mary E.  
APPLICANT: Utans, Ulrike  
TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/361,441B  
FILING DATE: 21-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/171,385  
FILING DATE: 21-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/014001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 295 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-361-441B-16  
Query Match 1.0%; Score 20; DB 3; Length 295;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2013 ATAGAGAAAAA 2032  
Db 27 ATAGAGAAAAA 8

RESULT 9  
US-08-171-385-12/c  
Sequence 12, Application US/08171385  
Patent No. 5527884  
GENERAL INFORMATION:  
APPLICANT: Mary E. Russell  
APPLICANT: Ulrike Utans  
TITLE OF INVENTION: Mediators of Chronic Allograft  
TITLE OF INVENTION: Rejection  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,385  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/006001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-171-385-12  
Query Match 1.0%; Score 20; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2013 ATAGAGAAAAA 2032  
Db 35 ATAGAGAAAAA 16

RESULT 10  
US-08-361-441B-12/c  
Sequence 12, Application US/08361441B  
Patent No. 6077948  
GENERAL INFORMATION:  
APPLICANT: Russell, Mary E.  
APPLICANT: Utans, Ulrike  
TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/361,441B  
FILING DATE: 21-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/171,385  
FILING DATE: 21-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/014001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

```

: TOPOLOGY: linear
US-08-361-441B-12
:
: Query Match
: Best Local Similarity 100.0%; Score 20; DB 3; Length 362;
: Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2013 ATAGAGAAAAA 2032
:
: Db 35 ATAGAGAAAAA 16
:
RESULT 11
US-09-247-155-152
: Sequence 152; Application US/09247155A
: Patent No. 6312822
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, Jean-Baptiste
: APPLICANT: Duclert, Aymeric
: APPLICANT: Bouquellet, Lydie
: TITLE OF INVENTION: Complementary DNAs
: FILE REFERENCE: GENSET 021A
: CURRENT APPLICATION NUMBER: US/09/247,155A
: CURRENT FILING DATE: 1999-02-09
: EARLIER APPLICATION NUMBER: 60/074,121
: EARLIER FILING DATE: 1998-02-09
: EARLIER APPLICATION NUMBER: 60/081,563
: EARLIER FILING DATE: 1998-04-13
: EARLIER APPLICATION NUMBER: 60/096,116
: EARLIER FILING DATE: 1998-08-10
: EARLIER APPLICATION NUMBER: 60/099,273
: EARLIER FILING DATE: 1998-10-04
: NUMBER OF SEQ ID NOS: 182
: SOFTWARE: Patent.pm
: SEQ ID NO 152
: LENGTH: 387
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 70..321
: FEATURE:
: NAME/KEY: sig.peptide
: LOCATION: 70..234
: OTHER INFORMATION: Von Heijne matrix
: OTHER INFORMATION: score 4.03999990463257
: OTHER INFORMATION: seq AVCAALLASHPTA/EV
: FEATURE:
: NAME/KEY: polyA.signal
: LOCATION: 364..369
: FEATURE:
: NAME/KEY: polyA-site
: LOCATION: 375..387
: US-09-247-155-152
:
Query Match
: Best Local Similarity 100.0%; Score 20; DB 4; Length 387;
: Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2011 AATAGAGAAAAA 2030
:
: Db 368 AATAGAGAAAAA 387
:
RESULT 12
US-09-336-536-26
: Sequence 26; Application US/09336536
: Patent No. 6406884
: GENERAL INFORMATION:
: APPLICANT: Leiby, K.
: APPLICANT: McKay, C.
: APPLICANT: Bossone, S.
: TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
```

```

: FILE REFERENCE: 7853-144
: CURRENT APPLICATION NUMBER: US/09/336,536
: CURRENT FILING DATE: 1999-06-18
: NUMBER OF SEQ ID NOS: 75
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 26
: LENGTH: 1869
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: all "n" positions
: OTHER INFORMATION: n=a, c, g, or t
US-09-336-536-26
:
Query Match
: Best Local Similarity 100.0%; Score 20; DB 4; Length 1869;
: Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2013 ATAGAGAAAAA 2032
:
: Db 1840 ATAGAGAAAAA 1859
:
RESULT 13
US-09-360-545-68
: Sequence 68; Application US/09360545
: Patent No. 6429014
: GENERAL INFORMATION:
: APPLICANT: Croteau, Rodney B
: APPLICANT: Bohmann, Jorg
: APPLICANT: Steele, Christopher L
: APPLICANT: Phillips, Michael A
: TITLE OF INVENTION: MONOTERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
: FILE REFERENCE: wslu13885
: CURRENT APPLICATION NUMBER: US/09/360,545
: CURRENT FILING DATE: 1999-07-26
: EARLIER APPLICATION NUMBER: 60/052,249
: EARLIER FILING DATE: 1997-11-07
: EARLIER APPLICATION NUMBER: PCT/US98/14528
: EARLIER FILING DATE: 1998-07-10
: NUMBER OF SEQ ID NOS: 107
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 68
: LENGTH: 2429
: TYPE: DNA
: ORGANISM: Abies grandis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (35)..(1945)
US-09-360-545-68
:
Query Match
: Best Local Similarity 100.0%; Score 20; DB 4; Length 2429;
: Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2013 ATAGAGAAAAA 2032
:
: Db 2407 ATAGAGAAAAA 2426
:
RESULT 14
US-09-813-133A-3
: Sequence 3; Application US/09813133A
: Patent No. 6455294
: GENERAL INFORMATION:
: APPLICANT: Gan, Weiniu et al
: TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
: FILE REFERENCE: C1001173
: CURRENT APPLICATION NUMBER: US/09/813,133A
: CURRENT FILING DATE: 2001-06-06
```

Mon Jan 13 09:13:56 2003

us-09-816-825-1.oli20.rni

Page 7

NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 55827  
TYPE: DNA  
ORGANISM: Human  
US-09-813-133A-3

Query Match 1.0%; Score 20; DB 4; Length 55827;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AAATAGAAAAAAAAAAAAA 2030  
|||||  
Db 24049 AAATAGAAAAAAAAAAAAA 24068

Search completed: January 10, 2003, 21:40:17  
Job time : 205 secs

**This Page Blank (uspto)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 19:13:52 ; Search time 95 Seconds

(without alignments)  
9413.579 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032

Sequence: 1 ggctcagagccagatgctc.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO\_NUC Gapop 60.0 , Gapext 60.0

Searched: 389086 seqs, 220051671 residues

Word size : 20

Total number of hits satisfying chosen parameters: 47

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Published.Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2032	100.0	2032	9	US-10-007-262-2
2	2032	100.0	2032	10	US-09-816-825-1
3	505	24.9	505	10	US-09-998-598-2595
4	505	24.9	517	10	US-09-919-580-194
5	286	14.1	389	10	US-09-919-580-264
6	59	2.9	2544	10	US-09-927-602-1
7	59	2.9	48436	10	US-09-927-602-38
8	37	1.8	37	9	US-10-007-262-7
9	37	1.8	37	10	US-09-816-825-7
10	22	1.1	1782	10	US-09-917-800A-1462
11	22	1.1	2557	10	US-09-812-102-10
12	21	1.0	242	10	US-09-960-352-3651
13	21	1.0	260	10	US-09-983-965-5796
14	21	1.0	280	10	US-09-960-352-4352
15	21	1.0	411	9	US-09-954-531-265
16	21	1.0	411	10	US-09-880-107-1438
17	21	1.0	785	10	US-09-770-445-859
18	21	1.0	986	10	US-09-770-445-263
19	21	1.0	998	10	US-09-925-301-297

20	21	1.0	1263	9	US-10-016-157A-89	Sequence 89, Appl
21	21	1.0	1540	10	US-09-998-598-352	Sequence 352, App
22	21	1.0	1469	10	US-09-919-770-1	Sequence 1, Appl
23	21	1.0	1792	10	US-09-925-301-259	Sequence 259, App
24	21	1.0	2000	9	US-09-938-842A-4722	Sequence 4722, Ap
25	21	1.0	2000	9	US-09-925-301-572	Sequence 572, Ap
26	20	1.0	101	10	US-09-983-965-3358	Sequence 3358, App
27	20	1.0	161	10	US-09-867-701-9309	Sequence 9309, Ap
28	20	1.0	176	10	US-09-983-965-3866	Sequence 2866, Ap
29	20	1.0	326	10	US-09-960-352-6131	Sequence 8131, Ap
30	20	1.0	394	10	US-09-960-352-5062	Sequence 5062, Ap
31	20	1.0	398	10	US-09-960-352-11099	Sequence 11099, A
32	20	1.0	403	10	US-09-960-352-7945	Sequence 7945, Ap
33	20	1.0	410	10	US-09-919-580-294	Sequence 294, App
34	20	1.0	430	10	US-09-960-352-5747	Sequence 9747, App
35	20	1.0	474	10	US-09-834-975-616	Sequence 616, App
36	20	1.0	477	9	US-09-938-842A-3890	Sequence 3890, Ap
37	20	1.0	498	10	US-09-764-846-16	Sequence 16, Appl
38	20	1.0	517	10	US-09-917-800A-306	Sequence 306, App
39	20	1.0	554	10	US-09-919-580-462	Sequence 462, App
40	20	1.0	615	10	US-09-867-550-1485	Sequence 1485, App
41	20	1.0	630	10	US-09-925-300-548	Sequence 548, App
42	20	1.0	1117	10	US-09-925-302-240	Sequence 240, App
43	20	1.0	1389	10	US-09-820-893-14	Sequence 14, Appl
44	20	1.0	1965	9	US-09-938-842A-4048	Sequence 4048, Ap
45	20	1.0	2429	9	US-10-041-007-27	Sequence 27, Appl
46	20	1.0	2441	10	US-09-822-830A-422	Sequence 422, App
47	20	1.0	3492	10	US-09-529-063-78	Sequence 78, Appl

#### ALIGNMENTS

RESULT 1  
US-10-007-262-2  
Sequence 2, Appl  
Patent No. US20020164748A1  
GENERAL INFORMATION:  
APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Tangemann, Kirsten  
APPLICANT: Hemmerich, Stefan  
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107CIP  
CURRENT APPLICATION NUMBER: US/10/007,262  
PRIOR FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 2032  
TYPE: DNA  
ORGANISM: H. sapiens  
US-10-007-262-2  
Query Match 100.0%; Score 2032; DB 9; Length 2032;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 89, Appl  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 352, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 1, Appl  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 259, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 4722, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 572, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 3358, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 9309, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 2866, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 8131, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 5062, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 11099, A  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 7945, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 294, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 9747, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 616, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 3890, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 16, Appl  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 306, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 462, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 1485, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 548, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 240, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 14, Appl  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 4048, Ap  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 27, Appl  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 422, App  
1 GGCTCAGGCGCAGAGGCCCTCCAGCTCTGGGGGAAATGCTTCTCACTTTGCTTCCAG 60  
Sequence 78, Appl

181 AATCTACTGCTTAAAAAATGAGCTCCGCTGTTCTGTTCCAGATGAGCACTT 240  
241 GGCCTATTTCTTCACATGTCACAGCCACACATCAGCTCCCTGTCATGAGGACAGCC 300  
241 GGCCTATTTCTTCACATGTCACAGCCACACATCAGCTCCCTGTCATGAGGACAGCC 300  
301 CGAGCGCATGCACTGCTGTTCTGTTCTGCTGCGCTGCTGCTCTTCTTTGAGGACA 360  
301 CGAGCGCATGCACTGCTGTTCTGTTCTGCTGCGCTGCTGCTCTTCTTTGAGGACA 360  
361 GCTTTTGGGAGAGCCACAGATGTTTCTACCTGATGAGACCCGCGCTGAGACGATGAT 420  
361 GCTTTTGGGAGAGCCACAGATGTTTCTACCTGATGAGACCCGCGCTGAGACGATGAT 420  
421 GACCTTCAAGCAGACAGCCGCTGATGCTGACATGCTGCTGCTGCTGCTGCTGCTGCT 480  
421 GACCTTCAAGCAGACAGCCGCTGATGCTGACATGCTGCTGCTGCTGCTGCTGCTGCT 480  
481 CGTCTCTGTTGGACATGAGCGCTTGTGATGCTTACATGGAACCTGCTGCTGCTGCTGCT 540  
481 CGTCTCTGTTGGACATGAGCGCTTGTGATGCTTACATGGAACCTGCTGCTGCTGCTGCT 540  
541 GTCCAGCCTCTTCTAGTGGAGAAACAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
541 GTCCAGCCTCTTCTAGTGGAGAAACAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
601 CATCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
601 CATCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
661 CTTGAGGTGTTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
661 CTTGAGGTGTTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
721 CTTTCTCAACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
721 CTTTCTCAACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
781 CGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
781 CGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
841 TCTCATGATTTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
841 TCTCATGATTTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
901 CCAACCTTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
901 CCAACCTTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
961 CCACTGCTTGGCCAAAGCCCTGAGAAAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
961 CCACTGCTTGGCCAAAGCCCTGAGAAAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
1021 TCGAGCCCCCTGAGCCAGACTTCCGAATGATGATGATGATGATGATGATGATGATGAT 1080  
1021 TCGAGCCCCCTGAGCCAGACTTCCGAATGATGATGATGATGATGATGATGATGATGAT 1080  
1081 CCACTCTCAGACCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
1081 CCACTCTCAGACCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
1141 CCAACCAATGCAAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
1141 CCAACCAATGCAAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
1201 TGAAGAAGTTTCTGACATTTGAGAAAGCTGTTGGGATGATGATGATGATGATGATGAT 1260  
1201 TGAAGAAGTTTCTGACATTTGAGAAAGCTGTTGGGATGATGATGATGATGATGATGAT 1260  
1261 CCAAGTCAGATCTGACAAAGAGAGAAAGCTGTTGCTGATCTTCTGCTGCTGCTGCTGCT 1320  
1261 CCAAGTCAGATCTGACAAAGAGAGAAAGCTGTTGCTGATCTTCTGCTGCTGCTGCTGCT 1320

1261 CCAAGTCAGATCTGACAAAGAGAGAAAGCTGTTGCTGATCTTCTGCTGCTGCTGCTGCT 1320  
1321 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTTGTGCTGCTGCTGCTGCTGCT 1380  
1321 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTTGTGCTGCTGCTGCTGCTGCT 1380  
1381 TCACTACTTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
1381 TCACTACTTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
1441 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
1441 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
1501 GTGTCCATGCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1501 GTGTCCATGCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1561 CAGTGAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
1561 CAGTGAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
1621 AGACTTGTGCTGAGAGGCTATTAGCAGCAGACAGATGATGATGATGATGATGATGATGAT 1680  
1621 AGACTTGTGCTGAGAGGCTATTAGCAGCAGACAGATGATGATGATGATGATGATGATGAT 1680  
1681 ACCTCCCTGCTCAGATCTTGGCCATGAGGATGATGATGATGATGATGATGATGATGATGAT 1740  
1681 ACCTCCCTGCTCAGATCTTGGCCATGAGGATGATGATGATGATGATGATGATGATGATGAT 1740  
1741 ATTTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
1741 ATTTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
1801 GTGGAAACAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
1801 GTGGAAACAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
1861 AATATGAACAAATCTGCTGCAACAAAGAGCAAGCTTAAATGATGATGATGATGATGATGAT 1920  
1861 AATATGAACAAATCTGCTGCAACAAAGAGCAAGCTTAAATGATGATGATGATGATGATGAT 1920  
1921 TGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
1921 TGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
1981 AAGCTGCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2032  
1981 AAGCTGCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2032

RESULT 2  
US-09-816-825-1  
Sequence 1, Application US/09816825  
Patent No. US20010051370A1  
GENERAL INFORMATION:  
APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Hemmerich, Stefan  
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107CON  
CURRENT APPLICATION NUMBER: US/09/816,825  
CURRENT FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/045,284  
PRIOR FILING DATE: 1998-03-20  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 1  
LENGTH: 2032  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-816-825-1



```

; APPLICANT: Stolck, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2595
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2595

Query Match
Best Local Similarity 100.0%; Pred. No. 4e-204;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1355 GCGTTGTGCGCAGCTGCTGTCAGCCTCAGTCTTCTGTGAATGCTTTGAGCCCTGC 1414
Db 1 GGGTTTGGTGCACCTGCTGTCAGCCTCAGTCTTCTGTGAATGCTTTGAGCCCTGC 60
QY 1415 CTACATCTCTGAGCCTTAACATCATGCTGTGGGTATCAGACTGAGTGTGCTGTC 1474
Db 61 CTACATCTCTGAGCCTTAACATCATGCTGTGGGTATCAGACTGAGTGTGCTGTC 120
QY 1475 ACACGTGCTCAGCAGAGAGCACTTTTGTGCATGCTTGTCTAGAAAACAGACTGGG 1534
Db 121 ACACGTGCTCAGCAGAGAGCACTTTTGTGCATGCTTGTCTAGAAAACAGACTGGG 180
QY 1535 AACCTTATGTAGCAGACATCCACAGAGAAACAGGGATTTGCTTCTTTCTT 1594
Db 181 AACCTTATGTAGCAGACATCCACAGAGAAACAGGGATTTGCTTCTTTCTT 240
QY 1595 GATCTTCTCTCTGCGCAGACTTCAGAGACTTTGTGGCTGAGGCGCTATTAAAGCAGAC 1654
Db 241 GATCTTCTCTCTGCGCAGACTTCAGAGACTTTGTGGCTGAGGCGCTATTAAAGCAGAC 300
QY 1655 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGGATG 1714
Db 301 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGGATG 360
QY 1715 GATCTTTCACCAAGAGCTCACAGCACTTTTCCAGAGATGCAAAATTCGAGCCCTTG 1774
Db 361 GATCTTTCACCAAGAGCTCACAGCACTTTTCCAGAGATGCAAAATTCGAGCCCTTG 420
QY 1775 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGGTGATGCTTATGAGCTT 1834
Db 421 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGGTGATGCTTATGAGCTT 480
QY 1835 GACCATCAGACGATCGGTATTCAG 1859
Db 481 GACCATCAGACGATCGGTATTCAG 505

RESULT 4
US-09-919-580-194
; Sequence 194, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 194
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-194

Query Match
Best Local Similarity 100.0%; Pred. No. 4e-204;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1355 GCGTTGTGCGCAGCTGCTGTCAGCCTCAGTCTTCTGTGAATGCTTTGAGCCCTGC 1414
Db 13 GGGTTTGGTGCACCTGCTGTCAGCCTCAGTCTTCTGTGAATGCTTTGAGCCCTGC 72
QY 1415 CTACATCTCTGAGCCTTAACATCATGCTGTGGGTATCAGACTGAGTGTGCTGTC 1474
Db 73 CTACATCTCTGAGCCTTAACATCATGCTGTGGGTATCAGACTGAGTGTGCTGTC 132
QY 1475 ACACGTGCTCAGCAGAGAGCACTTTTGTGCATGCTTGTCTAGAAAACAGACTGGG 1534
Db 133 ACACGTGCTCAGCAGAGAGAGCACTTTTGTGCATGCTTGTCTAGAAAACAGACTGGG 192
QY 1535 AACCTTATGTAGCAGACATCCACAGAGAAACAGGGATTTGCTTCTTTCTT 1594
Db 193 AACCTTATGTAGCAGACATCCACAGAGAAACAGGGATTTGCTTCTTTCTT 252
QY 1595 GATCTTCTCTCTGCGCAGACTTCAGAGACTTTGTGGCTGAGGCGCTATTAAAGCAGAC 1654
Db 253 GATCTTCTCTCTGCGCAGACTTCAGAGACTTTGTGGCTGAGGCGCTATTAAAGCAGAC 312
QY 1655 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGGATG 1714
Db 313 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGGATG 372
QY 1715 GATCTTTCACCAAGAGCTCACAGCACTTTTCCAGAGATGCAAAATTCGAGCCCTTG 1774
Db 373 GATCTTTCACCAAGAGCTCACAGCACTTTTCCAGAGATGCAAAATTCGAGCCCTTG 432
QY 1775 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGGTGATGCTTATGAGCTT 1834
Db 433 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGGTGATGCTTATGAGCTT 492
QY 1835 GACCATCAGACGATCGGTATTCAG 1859
Db 493 GACCATCAGACGATCGGTATTCAG 517

RESULT 5
US-09-919-580-264
; Sequence 264, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-264

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e-111;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1732 CTCACACAGATTTTCACAGAGATGCAAAATTCAGCCCTTGAGGCTTCCAGTGATTC 1791
```



```
RESULT 10
; Sequence 1462, Application us/09917800A
; Patent No. US20020113462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1462
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U39208
US-09-917-800A-1462

Query Match          1.1%; Score 22; DB 10; Length 1782;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATTAAGAGAAAAA 2032
Db 1756 AATTAAGAGAAAAA 1777

RESULT 11
; Sequence 10, Application US/09812102
; Patent No. US20020055179A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E
; TITLE OF INVENTION: No. US20020055179A1el G-Protein Coupled Receptor Homologs
; FILE REFERENCE: 5800-41 035800/183478
; CURRENT APPLICATION NUMBER: US/09/812,102
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/364,769
; PRIOR FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 2557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: GPCR-SECRETIN
; NAME/KEY: misc_feature
; LOCATION: (1)..(2557)
; OTHER INFORMATION: n = a, t, c, or g
US-09-812-102-10

Query Match          1.1%; Score 22; DB 10; Length 2557;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATTAAGAGAAAAA 2032
Db 2500 AATTAAGAGAAAAA 2521

RESULT 12
; Sequence 3651, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511-006/37-21(102981C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3651
; LENGTH: 242
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 16-LIB3058-047-Q1-K1-D11
US-09-960-352-3651

Query Match          1.0%; Score 21; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATTAAGAGAAAAA 2032
Db 57 AATTAAGAGAAAAA 37

RESULT 13
; Sequence 5796, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(102971C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5796
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 59-LIB34-047-Q1-E1-G12
US-09-983-965-5796
```

Mon Jan 13 09:13:57 2003

us-09-816-825-1.01i20.rmpb

Page 7

Query Match 1.0%; Score 21; DB 10; Length 260;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATAAGAAAAAAAAAAAAA 2031  
|||||  
DB 207 AATAAGAAAAAAAAAAAAA 227

RESULT 14  
US-09-960-352-4352/c  
Sequence 4352, Application US/09560352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 4352  
LENGTH: 280  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 19-LIB3058-003-Q1-K1-E3  
US-09-960-352-4352

Query Match 1.0%; Score 21; DB 10; Length 280;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAAAAAAAAAA 2032  
|||||  
DB 49 AATAAGAAAAAAAAAAAAA 29

RESULT 15  
US-09-954-531-265/c  
Sequence 265, Application US/0954531  
Patent No. US20020165180A1  
GENERAL INFORMATION:  
APPLICANT: Weaver, Zoe  
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
FILE REFERENCE: 689290-77  
CURRENT APPLICATION NUMBER: US/09/954,531  
CURRENT FILING DATE: 2002-05-02  
PRIOR APPLICATION NUMBER: US/60/233,133  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/234,009  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,034  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,509  
PRIOR FILING DATE: 2000-09-22  
PRIOR APPLICATION NUMBER: US/60/234,567  
PRIOR FILING DATE: 2000-09-22  
NUMBER OF SEQ ID NOS: 1392  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 265  
LENGTH: 411  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-954-531-265

Query Match 1.0%; Score 21; DB 9; Length 411;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAAAAAAAAAA 2032  
|||||  
DB 44 AATAAGAAAAAAAAAAAAA 24

RESULT 16  
US-09-880-107-1438/c  
Sequence 1438, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1438  
LENGTH: 411  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA598749  
US-09-880-107-1438

Query Match 1.0%; Score 21; DB 10; Length 411;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAAAAAAAAAA 2032  
|||||  
DB 44 AATAAGAAAAAAAAAAAAA 24

RESULT 17  
US-09-770-445-859/c  
Sequence 859, Application US/09770445  
Patent No. US2002023281A1  
GENERAL INFORMATION:  
APPLICANT: Gerlach, Jorn  
APPLICANT: An, Yong-Olang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yi, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Krickler, Maya  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2023US (PARA-012PRV)  
CURRENT APPLICATION NUMBER: US/09/770,445  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: US 60/178,472  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999

Query Match 1.0%; Score 21; DB 9; Length 411;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 859  
 LENGTH: 785  
 TYPE: DNA  
 ORGANISM: Arabidopsis thaliana  
 US-09-770-445-859

Query Match  
 Best Local Similarity 100.0%; Score 21; DB 10; Length 785;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAGAAAAA 2032  
 |||||  
 Db 31 AATAGAGAAAAA 11

RESULT 18  
 US-09-770-445-263  
 Sequence 263, Application US/09770445  
 Patent No. US20020023281A1  
 GENERAL INFORMATION:

APPLICANT: Gorlach, Jörn  
 APPLICANT: An, Jong-Qiang  
 APPLICANT: Hamilton, Carol M.  
 APPLICANT: Price, Jennifer L.  
 APPLICANT: Raines, Tracy M.  
 APPLICANT: Yu, Yang  
 APPLICANT: Rameaka, Joshua G.  
 APPLICANT: Page, Amy  
 APPLICANT: Matthew, Abraham V.  
 APPLICANT: Ledford, Brooke L.  
 APPLICANT: Moessner, Jeffrey P.  
 APPLICANT: Haas, William David  
 APPLICANT: Garcia, Carlos A.  
 APPLICANT: Krieger, Maja  
 APPLICANT: Slader, Ted  
 APPLICANT: Davis, Keith R.  
 APPLICANT: Allen, Keith  
 APPLICANT: Hoffman, Neil  
 APPLICANT: Hurban, Patrick  
 TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
 FILE REFERENCE: 2023US (PARA-012PRV)  
 CURRENT APPLICATION NUMBER: US/09/770,445  
 CURRENT FILING DATE: 2001-01-26  
 PRIOR APPLICATION NUMBER: US 60/178,472  
 PRIOR FILING DATE: 2000-01-27  
 NUMBER OF SEQ ID NOS: 999  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 263  
 LENGTH: 986  
 TYPE: DNA  
 ORGANISM: Arabidopsis thaliana  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)-(986)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-770-445-263

Query Match  
 Best Local Similarity 100.0%; Score 21; DB 10; Length 986;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAGAAAAA 2032  
 |||||  
 Db 158 AATAGAGAAAAA 178

RESULT 19  
 US-09-925-301-297  
 Sequence 297, Application US/09925301  
 Patent No. US20020052308A1  
 GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA106  
 CURRENT APPLICATION NUMBER: US/09/925,301  
 CURRENT FILING DATE: 2001-08-10  
 PRIOR APPLICATION NUMBER: PCT/US00/05882  
 PRIOR FILING DATE: 2000-03-08  
 PRIOR APPLICATION NUMBER: 60/124,270  
 PRIOR FILING DATE: 1999-03-12  
 NUMBER OF SEQ ID NOS: 1694  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 297  
 LENGTH: 998  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-925-301-297

Query Match  
 Best Local Similarity 100.0%; Score 21; DB 10; Length 998;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAGAAAAA 2032  
 |||||  
 Db 944 AATAGAGAAAAA 964

RESULT 20  
 US-10-016-157A-89  
 Sequence 89, Application US/10016157A  
 Publication No. US20020192220A1  
 GENERAL INFORMATION:  
 APPLICANT: Sun, Yonggang  
 APPLICANT: Recipon, Heive  
 APPLICANT: Ghosh, Malayika  
 APPLICANT: Liu, Chenghua  
 TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and  
 FILE REFERENCE: DEX-0253  
 CURRENT APPLICATION NUMBER: US/10/016,157A  
 CURRENT FILING DATE: 2001-10-31  
 PRIOR APPLICATION NUMBER: US 60/244,717  
 PRIOR FILING DATE: 2000-10-31  
 NUMBER OF SEQ ID NOS: 250  
 SOFTWARE: PatentIn Version 3.1  
 SEQ ID NO: 89  
 LENGTH: 1263  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-016-157A-89

Query Match  
 Best Local Similarity 100.0%; Score 21; DB 9; Length 1263;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATAGAGAAAAA 2031  
 |||||  
 Db 1243 AATAGAGAAAAA 1263

RESULT 21  
 US-09-998-598-352  
 Sequence 352, Application US/09998598  
 Patent No. US20020150922A1  
 GENERAL INFORMATION:  
 APPLICANT: Stolk, John A.  
 APPLICANT: Xu, Jiangchun  
 APPLICANT: Chenaault, Ruth A.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 FILE REFERENCE: 210121.361  
 CURRENT APPLICATION NUMBER: US/09/998,598  
 CURRENT FILING DATE: 2001-11-16  
 NUMBER OF SEQ ID NOS: 2606



SOFTWARE: Corixa Invention Disclosure Database  
SEQ ID NO 352  
LENGTH: 1340  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-998-598-352

Query Match 1.0%; Score 21; DB 10; Length 1340;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032  
Db 1316 AATAAGAAAAA 1336

RESULT 22  
US-09-919-770-1  
Sequence 1, Application US/09919770  
Patent No. US20020048577A1  
GENERAL INFORMATION:  
APPLICANT: Bornstein, Paul  
APPLICANT: Kyriakides, Themis  
APPLICANT: Ratner, Buddy  
APPLICANT: Giachelli, Cecilia  
APPLICANT: Martinson, Laura  
APPLICANT: Scatena, Maria  
TITLE OF INVENTION: Methods and Devices to Modulate the Wound Response  
FILE REFERENCE: USFW17618  
CURRENT APPLICATION NUMBER: US/09/919,770  
CURRENT FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: US 60/222,071  
PRIOR FILING DATE: 2000-08-01  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 1469  
TYPE: DNA  
ORGANISM: Homo Sapien  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (102)..(1001)  
US-09-919-770-1

Query Match 1.0%; Score 21; DB 10; Length 1469;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032  
Db 1435 AATAAGAAAAA 1455

RESULT 23  
US-09-925-301-259  
Sequence 259, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 259  
LENGTH: 1792  
TYPE: DNA  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (107)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (487)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1306)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-301-259

Query Match 1.0%; Score 21; DB 10; Length 1792;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032  
Db 1768 AATAAGAAAAA 1768

RESULT 24  
US-09-938-842A-4722  
Sequence 4722, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
TITLE OF INVENTION: SAME, AND METHODS OF USE  
FILE REFERENCE: SCRIPI300.3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 4722  
LENGTH: 2000  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4722

Query Match 1.0%; Score 21; DB 9; Length 2000;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032  
Db 1735 AATAAGAAAAA 1755

RESULT 25  
US-09-925-301-572  
Sequence 572, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 572  
LENGTH: 2013  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-301-572

Query Match 1.0%; Score 21; DB 10; Length 2013;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAAAAA 2032  
Db 1993 AATAGAAAAA 2013

RESULT 26  
US-09-983-965-3358/c  
Sequence 3358, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengding

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 37-21(10297)C

CURRENT APPLICATION NUMBER: US/09/983,965

PRIOR FILING DATE: 2001-10-26

PRIOR APPLICATION NUMBER: US 09/465,231

PRIOR FILING DATE: 1999-12-15

PRIOR APPLICATION NUMBER: US 60/113,678

PRIOR FILING DATE: 1998-12-17

NUMBER OF SEQ ID NOS: 5912

SEQ ID NO 3358

LENGTH: 101

TYPE: DNA

ORGANISM: Bos taurus

FEATURE:

OTHER INFORMATION: Clone ID: 33-LIB3058-031-Q1-K1-A2

US-09-983-965-3358

Query Match 1.0%; Score 20; DB 10; Length 101;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 AATAGAAAAA 2032  
Db 37 AATAGAAAAA 18

RESULT 27  
US-09-867-701-9309  
Sequence 9309, Application US/09867701  
Patent No. US20020132237A1  
GENERAL INFORMATION:

APPLICANT: Agiate, Paul A.

APPLICANT: Harlocker, Susan L.

APPLICANT: Jones, Robert

APPLICANT: Harlocker, Susan L.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

FILE REFERENCE: 210121.497

CURRENT APPLICATION NUMBER: US/09/867,701

PRIOR FILING DATE: 2001-05-29

NUMBER OF SEQ ID NOS: 10912

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9309

LENGTH: 161

TYPE: DNA

ORGANISM: Homo sapien

US-09-867-701-9309

Query Match 1.0%; Score 20; DB 10; Length 161;

Best Local Similarity 100.0%; Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 AATAGAAAAA 2032  
Db 111 AATAGAAAAA 130

RESULT 28  
US-09-983-965-2866/c  
Sequence 2866, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengding

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 37-21(10297)C

CURRENT APPLICATION NUMBER: US/09/983,965

PRIOR FILING DATE: 2001-10-26

PRIOR APPLICATION NUMBER: US 09/465,231

PRIOR FILING DATE: 1999-12-15

PRIOR APPLICATION NUMBER: US 60/113,678

PRIOR FILING DATE: 1998-12-17

NUMBER OF SEQ ID NOS: 5912

SEQ ID NO 2866

LENGTH: 176

TYPE: DNA

ORGANISM: Bos taurus

FEATURE:

OTHER INFORMATION: Clone ID: 17-LIB3056-004-Q1-K1-E1

US-09-983-965-2866

Query Match 1.0%; Score 20; DB 10; Length 176;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 AATAGAAAAA 2032  
Db 44 AATAGAAAAA 25

RESULT 29  
US-09-960-352-8131/c  
Sequence 8131, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengding

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

PRIOR FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 8131

LENGTH: 326

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 35-LIB3057-010-Q1-K1-A12

US-09-960-352-8131

Query Match 1.0%; Score 20; DB 10; Length 326;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 AATAGAAAAA 2032  
Db 42 AATAGAAAAA 23

```
RESULT 30
US-09-960-352-5062/C
; Sequence 5062, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: MUSCLE ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5062
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 22-LIB3058-008-Q1-K1-F5
US-09-960-352-5062

Query Match          1.0%; Score 20; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
Db 44 ATAGAGAAAAA 25

RESULT 31
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-O1-E1-D8
US-09-960-352-11099

Query Match          1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 GATGCCATCTTGCTCAT 248
Db 240 GATGCCATCTTGCTCAT 221

RESULT 32
US-09-960-352-7945/C
; Sequence 7945, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7945
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB3058-053-Q1-K1-A6
US-09-960-352-7945

Query Match          1.0%; Score 20; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
Db 41 ATAGAGAAAAA 22

RESULT 33
US-09-919-580-294/C
; Sequence 294, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 294
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 168..178, 195, 288, 291, 305..401
; OTHER INFORMATION: n = A,T,C or G
US-09-919-580-294

Query Match          1.0%; Score 20; DB 10; Length 410;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
Db 28 ATAGAGAAAAA 9

RESULT 34
US-09-960-352-9747/C
; Sequence 9747, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 9747
; LENGTH: 430
```

TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 42-LIB3057-018-Q1-K1-C6  
US-09-960-352-9747

Query Match  
Best Local Similarity 100.0%; Score 20; DB 10; Length 430;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 400 ATAGAGAAAAA 381

RESULT 35  
US-09-834-975-616  
Sequence 616, Application US/09834975  
Patent No. US20020110815A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Brown, Jeffrey  
APPLICANT: Bolt, Andrew  
APPLICANT: Van Hufel, Christophe  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
FILE REFERENCE: MRI-016B  
CURRENT APPLICATION NUMBER: US/09/834,975  
CURRENT FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/197,538  
PRIOR FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 1046  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 616  
LENGTH: 474  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: (1)..(474)  
OTHER INFORMATION: n = A,T,C or G  
US-09-834-975-616

Query Match  
Best Local Similarity 100.0%; Score 20; DB 10; Length 474;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 214 ATAGAGAAAAA 233

RESULT 36  
US-09-938-842A-3890/C  
Sequence 3890, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kieps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 3890  
LENGTH: 477  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3890

Query Match  
Best Local Similarity 100.0%; Score 20; DB 9; Length 477;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 85 ATAGAGAAAAA 66

RESULT 37  
US-09-764-846-16  
Sequence 16, Application US/09764846  
Patent No. US20020102638A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT212  
CURRENT APPLICATION NUMBER: US/09/764,846  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper.  
NUMBER OF SEQ ID NOS: 348  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 16  
LENGTH: 498  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-846-16

Query Match  
Best Local Similarity 100.0%; Score 20; DB 10; Length 498;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 460 ATAGAGAAAAA 479

RESULT 38  
US-09-917-800A-306/C  
Sequence 306, Application US/09917800A  
Patent No. US20020119462A1  
GENERAL INFORMATION:  
APPLICANT: Mendrick, Donna  
APPLICANT: Porter, Mark  
APPLICANT: Johnson, Kory  
APPLICANT: Castle, Arthur  
APPLICANT: Elashoff, Michael  
TITLE OF INVENTION: Molecular Toxicology Modeling  
FILE REFERENCE: 44921-5038-US  
CURRENT APPLICATION NUMBER: US/09/917,800A  
CURRENT FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: US 60/222,040  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: US 60/222,880  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: US 60/290,029  
PRIOR FILING DATE: 2001-05-11  
PRIOR APPLICATION NUMBER: US 60/290,645  
PRIOR FILING DATE: 2001-05-15  
PRIOR APPLICATION NUMBER: US 60/292,336  
PRIOR FILING DATE: 2001-05-22  
PRIOR APPLICATION NUMBER: US 60/295,798  
PRIOR FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: US 60/297,457  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,884

;; PRIOR FILING DATE: 2001-06-19  
;; PRIOR APPLICATION NUMBER: US 60/303,459  
;; PRIOR FILING DATE: 2001-07-09  
;; NUMBER OF SEQ ID NOS: 1740  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 306  
;; LENGTH: 517  
;; TYPE: DNA  
;; ORGANISM: Rattus norvegicus  
;; FEATURE:  
;; OTHER INFORMATION: Genbank Accession No. US20020119462A1,AA924767  
US-09-917-800A-306

Query Match 1.0%; Score 20; DB 10; Length 517;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
|||||

DB 22 ATAGAGAAAAA 3

RESULT 39  
US-09-919-580-462/C  
; Sequence 462, Application US/09919580  
; Patent No. US20020110832A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121,552  
; CURRENT APPLICATION NUMBER: US/09/919,580  
; CURRENT FILING DATE: 2001-07-30  
; NUMBER OF SEQ ID NOS: 934  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 462  
; LENGTH: 554  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 38, 169, 229, 278, 289, 292, 312, 329, 340, 368, 374, 414,  
; LOCATION: 513, 537  
; OTHER INFORMATION: n = A,T,C or G  
US-09-919-580-462

Query Match 1.0%; Score 20; DB 10; Length 554;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
|||||

DB 29 ATAGAGAAAAA 10

RESULT 40  
US-09-867-550-1485/C  
; Sequence 1485, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehrtan, Fuad  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427

;; PRIOR FILING DATE: 2000-05-30  
;; NUMBER OF SEQ ID NOS: 2125  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1485  
;; LENGTH: 615  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-867-550-1485

Query Match 1.0%; Score 20; DB 10; Length 615;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
|||||

DB 22 ATAGAGAAAAA 3

RESULT 41  
US-09-925-300-548  
; Sequence 548, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 548  
; LENGTH: 630  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (61)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-300-548

Query Match 1.0%; Score 20; DB 10; Length 630;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
|||||

DB 577 ATAGAGAAAAA 596

RESULT 42  
US-09-925-302-240  
; Sequence 240, Application US/09925302  
; Patent No. US2002004941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 240  
; LENGTH: 1117  
; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1113)  
OTHER INFORMATION: n equals a.t.g, or c  
US-09-925-302-240

Query Match 1.0%; Score 20; DB 10; Length 1117;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 1088 ATAGAGAAAAA 1107

RESULT 43  
US-09-820-893-14  
Sequence 14, Application US/09820893  
Patent No. US20020076705A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 31 Human Secreted Proteins  
FILE REFERENCE: P2033P1  
CURRENT APPLICATION NUMBER: US/09/820,893  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 09/531,119  
PRIOR FILING DATE: 2000-03-20  
PRIOR APPLICATION NUMBER: 60/102,895  
PRIOR FILING DATE: 1998-10-02  
NUMBER OF SEQ ID NOS: 140  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 1389  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-820-893-14

Query Match 1.0%; Score 20; DB 10; Length 1389;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 1347 ATAGAGAAAAA 1366

RESULT 44  
US-09-938-842A-4048  
Sequence 4048, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SAME, AND METHODS OF USE  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 4048  
LENGTH: 1965  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4048

Query Match 1.0%; Score 20; DB 9; Length 1965;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 1538 ATAGAGAAAAA 1557

RESULT 45  
US-10-041-007-27  
Sequence 27, Application US/10041007  
Patent No. US20020164736A1  
GENERAL INFORMATION:  
APPLICANT: Matsuda, Seiichi P.T.  
APPLICANT: Schepmann, Hala G  
TITLE OF INVENTION: Ginkgo Biloba Levopimaradiene Synthase  
FILE REFERENCE: P02081US1  
CURRENT APPLICATION NUMBER: US/10/041,007  
CURRENT FILING DATE: 2002-01-07  
PRIOR APPLICATION NUMBER: US 60/259,881  
PRIOR FILING DATE: 2001-01-05  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 27  
LENGTH: 2429  
TYPE: DNA  
ORGANISM: Abies grandis  
US-10-041-007-27

Query Match 1.0%; Score 20; DB 9; Length 2429;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032  
DB 2407 ATAGAGAAAAA 2426

RESULT 46  
US-09-822-830A-422  
Sequence 422, Application US/09822830A  
Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulukota, Kamalakari  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,604  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 422  
LENGTH: 2441  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-422

Query Match 1.0%; Score 20; DB 10; Length 2441;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032

Db 1104 ATAGAAAAAAAAAAAA 1123

```

RESULT 47
US-09-529-063-78
; Sequence 78. Application US/09529063.
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIRICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529,063
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 3492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (6)..(53)
; NAME/KEY: mat_peptide
; LOCATION: (54)..(2117)
; NAME/KEY: CDS
; LOCATION: (6)..(2117)
US-09-529-063-78
    
```

```

Query Match 1.0%; Score 20; DB 10; Length 3492;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2013 ATAGAAAAAAAAAAAA 2032
Db 3470 ATAGAAAAAAAAAAAA 3489
    
```

Search completed: January 10, 2003, 21:43:58  
 Job time : 331 secs

This page Blank (just to)





QY 121 SSIFQWENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPEVYKACRSYSHVYLKEVR 180  
 CC and/or a beta1,3gnt antisenase nucleic acid molecule. L-selectin  
 CC sulfoftransferase-2 (LSST-2) also directs MECA-79 antigen expression.  
 CC Alternatively, the expression or activity of LSST-2 or its active  
 CC fragment can be reduced in combination with reducing the expression or  
 CC activity of beta1,3gnt. The method is useful for treating L-selectin  
 CC mediated conditions such as Crohn's disease and ulcerative colitis,  
 CC inflammatory disorders of the skin such as allergic contact dermatitis,  
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
 CC sequence represents human LSST-2.

QY 121 SSIFQWENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPEVYKACRSYSHVYLKEVR 180  
 Db 121 SSIFQWENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPEVYKACRSYSHVYLKEVR 180  
 QY 181 FENLQSLVYLLKDPSPNLNLIYLVDRPRAVFRSRETKGDMIDSRIYMGHEQKLKED 240  
 CC 181 FENLQSLVYLLKDPSPNLNLIYLVDRPRAVFRSRETKGDMIDSRIYMGHEQKLKED 240  
 Db 181 FENLQSLVYLLKDPSPNLNLIYLVDRPRAVFRSRETKGDMIDSRIYMGHEQKLKED 240  
 QY 241 QPYVWQVYCQSLDIYKTQSLPQALQERYLLVREDLARAQAQTSRMTEFVGLFELP 300  
 CC 241 QPYVWQVYCQSLDIYKTQSLPQALQERYLLVREDLARAQAQTSRMTEFVGLFELP 300  
 Db 241 QPYVWQVYCQSLDIYKTQSLPQALQERYLLVREDLARAQAQTSRMTEFVGLFELP 300  
 QY 301 HQTWVHNITRKGMDHAFHTNARDALNVSOAMRWSLPYEKVSRLQACGDAMNLLGYR 360  
 CC 301 HQTWVHNITRKGMDHAFHTNARDALNVSOAMRWSLPYEKVSRLQACGDAMNLLGYR 360  
 Db 301 HQTWVHNITRKGMDHAFHTNARDALNVSOAMRWSLPYEKVSRLQACGDAMNLLGYR 360  
 QY 361 HVRSQEQRNLLDLSTWVPEQIH 386  
 CC 361 HVRSQEQRNLLDLSTWVPEQIH 386  
 Db 361 HVRSQEQRNLLDLSTWVPEQIH 386

RESULT 2  
 ID AAO11274 standard; Protein: 380 AA.  
 XX AAO11274:  
 AC 12-MAR-2002 (first entry)  
 DT 12-MAR-2002 (first entry)  
 XX 12-MAR-2002 (first entry)  
 DE Human L-selectin sulfoftransferase-2 (LSST-2) protein.  
 XX Human L-selectin sulfoftransferase-2 (LSST-2) protein.  
 KW Human; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;  
 KW L-selectin; L-selectin sulfoftransferase-2; Crohn's disease; diabetes;  
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;  
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;  
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;  
 KW antinflammatory; antiproliferative; antidiabetic; dermatological;  
 KW antiallergic.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN WO200185177-A1.  
 PD 15-NOV-2001.  
 XX 15-NOV-2001.  
 PF 10-MAY-2001; 2001WO-US15452.  
 PR 11-MAY-2000; 2000US-0569320.  
 PA (BORN-) BURNHAM INST.  
 PI Fukuda M, Yeh J, Hiraoka N;  
 DR WPT; 2002-075226/10.  
 DR N-PSDB; AAS16947.  
 XX New enzyme, useful for modifying acceptor molecule, comprises an  
 PT isolated L-selectin sulfoftransferase-2 that directs expression of  
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or  
 PT intestinal G10NAC 6-sulfoftransferase  
 PS Claim 21; Fig 4; 98pp; English.  
 XX The present invention provides a method of modifying an acceptor molecule  
 CC by contacting the acceptor with an isolated  
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active  
 CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The  
 CC invention also provides a method of treating or preventing an  
 CC L-selectin-mediated condition by reducing the expression or activity of a  
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done  
 CC by administering to the subject an oligosaccharide L-selectin antagonist  
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by

CC administering antibody material that specifically binds beta1,3gnt,  
 CC and/or a beta1,3gnt antisenase nucleic acid molecule. L-selectin  
 CC sulfoftransferase-2 (LSST-2) also directs MECA-79 antigen expression.  
 CC Alternatively, the expression or activity of LSST-2 or its active  
 CC fragment can be reduced in combination with reducing the expression or  
 CC activity of beta1,3gnt. The method is useful for treating L-selectin  
 CC mediated conditions such as Crohn's disease and ulcerative colitis,  
 CC inflammatory disorders of the skin such as allergic contact dermatitis,  
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
 CC sequence represents human LSST-2.

QY 7 MKLLFLVSQMAIILAFPHMYSHNISLSKQAPERNHYLTSWRSQSFVQQLFGQP 66  
 CC 7 MKLLFLVSQMAIILAFPHMYSHNISLSKQAPERNHYLTSWRSQSFVQQLFGQP 66  
 Db 1 MKLLFLVSQMAIILAFPHMYSHNISLSKQAPERNHYLTSWRSQSFVQQLFGQP 60  
 QY 67 DVFYLMPEPAWVMTFKQSTAMKLHMAVRDLIAVFLCDMSVFDAYMEGPRQSSLFGM 126  
 CC 67 DVFYLMPEPAWVMTFKQSTAMKLHMAVRDLIAVFLCDMSVFDAYMEGPRQSSLFGM 126  
 Db 61 DVFYLMPEPAWVMTFKQSTAMKLHMAVRDLIAVFLCDMSVFDAYMEGPRQSSLFGM 120  
 QY 127 ENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPEVYKACRSYSHVYLKEVRPNILOS 186  
 CC 127 ENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPEVYKACRSYSHVYLKEVRPNILOS 186  
 Db 121 ENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPEVYKACRSYSHVYLKEVRPNILOS 180  
 QY 187 LYPILKDPSPNLNLIYLVDRPRAVFRSRETKGDMIDSRIYMGHEQKLKEDQPYVYM 246  
 CC 187 LYPILKDPSPNLNLIYLVDRPRAVFRSRETKGDMIDSRIYMGHEQKLKEDQPYVYM 246  
 Db 181 LYPILKDPSPNLNLIYLVDRPRAVFRSRETKGDMIDSRIYMGHEQKLKEDQPYVYM 240  
 QY 247 QVYCQSLDIYKTQSLPQALQERYLLVREDLARAQAQTSRMTEFVGLFELPPIHQTV 306  
 CC 247 QVYCQSLDIYKTQSLPQALQERYLLVREDLARAQAQTSRMTEFVGLFELPPIHQTV 306  
 Db 241 QVYCQSLDIYKTQSLPQALQERYLLVREDLARAQAQTSRMTEFVGLFELPPIHQTV 300  
 QY 307 HNTTRGKMGMDHAFHTNARDALNVSOAMRWSLPYEKVSRLQACGDAMNLLGYRHRSQ 366  
 CC 307 HNTTRGKMGMDHAFHTNARDALNVSOAMRWSLPYEKVSRLQACGDAMNLLGYRHRSQ 366  
 Db 301 HNTTRGKMGMDHAFHTNARDALNVSOAMRWSLPYEKVSRLQACGDAMNLLGYRHRSQ 360  
 QY 367 EQRNLLDLSTWVPEQIH 386  
 CC 367 EQRNLLDLSTWVPEQIH 386  
 Db 361 EQRNLLDLSTWVPEQIH 380

RESULT 3  
 ID AAM93309 standard; Protein: 386 AA.  
 XX AAM93309:  
 AC AAM93309;  
 DT 06-NOV-2001 (first entry)  
 DE Human polypeptide, SEQ ID NO: 2817.  
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.  
 XX Homo sapiens.  
 PN EPI130094-A2.  
 PD 05-SEP-2001.  
 XX 05-SEP-2001.  
 PF 07-JUL-2000; 2000EP-0114089.  
 PR 08-JUL-1999; 99QP-0194486.  
 PR 11-JAN-2000; 2000JP-0118774.  
 PR 02-MAY-2000; 2000JP-0183765.  
 PA (HELI-) HELIX RES INST.  
 XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 DR WPI: 2001-524255/58.  
 DR N-PSDB; AAK94229.

XX 830 Primers useful for synthesizing full length cDNA clones and their  
 PT use in genetic manipulation -

XX Claim 8: SEQ ID NO 2817; 1380bp + sequence listing; English.

PS The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a polypeptide  
 CC encoded by a full length human cDNA of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 386 AA:

Query Match 73.8%; Score 285; DB 22; Length 386;  
 Best Local Similarity 99.7%; Pred No. 5.3e-277;  
 Matches 385; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSOMAILALFFHMYSHNISLSNKAQPERMHVLYLSMRSGSEFVGQ 60  
 DB 1 MLTPKMKLLFLVSOMAILALFFHMYSHNISLSNKAQPERMHVLYLSMRSGSEFVGQ 60  
 QY 61 LFGQHDVFLMPFAHVMVTEKOSTAMLMHAYRDLTAVALCDMSVDAYMEPPRQ 120  
 DB 61 LFGQHDVFLMPFAHVMVTEKOSTAMLMHAYRDLTAVALCDMSVDAYMEPPRQ 120  
 QY 121 SSLFOWNSRALCSAPACDIIPQDEITIPRAHGRLLCSQOPFEVEKACRSYSHVYLKEVR 180  
 DB 121 SSLFOWNSRALCSAPACDIIPQDEITIPRAHGRLLCSQOPFEVEKACRSYSHVYLKEVR 180  
 QY 181 FFMLOSLYPLKRPSTNLHVIHVRPRAVFRSRETKGDLMDISITVAGCHQKAKKD 240  
 DB 181 FFMLOSLYPLKRPSTNLHVIHVRPRAVFRSRETKGDLMDISITVAGCHQKAKKD 240  
 QY 241 QPYVWQVICOQSOLEIYKTIQSLPKALQERYLLVRYEDJAPAPVQTSRMYEFVGLFP 300  
 DB 241 QPYVWQVICOQSOLEIYKTIQSLPKALQERYLLVRYEDJAPAPVQTSRMYEFVGLFP 300  
 QY 301 HLOTWVHNITRGKGDHAFHINARDALNVSOAMRSLPEYKVSRIQKACGDMNLGVR 360  
 DB 301 HLOTWVHNITRGKGDHAFHINARDALNVSOAMRSLPEYKVSRIQKACGDMNLGVR 360  
 QY 361 HVRSEQORNLILDLSTWVPEQIH 386  
 DB 361 HVRSEQORNLILDLSTWVPEQIH 386

RESULT 4

AA79219  
 ID AAY79219 standard; Protein: 386 AA.

XX AAY79219;

DT 19-JUN-2000 (first entry)

DE Human transferase TRNSFS-11.

XX Transferrase; TRNSFS-11; human; antitumor; cell proliferation;

KM gastrointestinal disorder; developmental disorder;

KW genetic disorder; neurological disorder; reproductive disorder;  
 smooth muscle disorder; immunological disorder; inflammation;

KM diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.  
 XX Homo sapiens.

Key	Location/Qualifiers
FT Modified-site 121	/note= "potential O-phosphorylation"
FT Modified-site 107	/note= "potential O-phosphorylation"
FT Modified-site 217	/note= "potential O-phosphorylation"
FT Modified-site 252	/note= "potential O-phosphorylation"
FT Modified-site 364	/note= "potential O-phosphorylation"
FT Modified-site 380	/note= "potential O-phosphorylation"
FT Modified-site 35	/note= "potential O-phosphorylation"
FT Modified-site 50	/note= "potential O-phosphorylation"
FT Modified-site 81	/note= "potential O-phosphorylation"
FT Modified-site 287	/note= "potential O-phosphorylation"
FT Modified-site 243	/note= "potential O-phosphorylation"
FT Modified-site 30	/note= "potential O-phosphorylation"
FT Modified-site 308	/note= "potential N-glycosylation"
FT Modified-site 329	/note= "potential N-glycosylation"
FT Domain 7..23	/note= "potential N-glycosylation"
FT	/note= "transmembrane domain"

PN WO200014251-A2.  
 PD 16-MAR-2000.  
 PF 09-SEP-1999; 99WO-US20989.

XX 10-SEP-1998; 98US-0150657.

PR 04-NOV-1998; 98US-0186779.

PR 11-MAY-1999; 99US-0133642.

PA (INCYTE) INCYTE PHARM INC.

PI Tang YF, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;

PI Hillman JL, Azimzai Y;

DR WPI: 2000-256996/22.

DR N-PSDB; AA294211.

XX Human transferase proteins useful for preventing, diagnosing and

PT treating cancers and developmental, gastrointestinal, genetic,

PT immunological, neurological, reproductive and smooth muscle disorders -

XX Claim 1; Page 90-91, 113pp; English.

XX The present sequence is that of human transferase TRNSFS-11, 1 of  
 CC 15 claimed novel human transferase proteins of the invention (see  
 CC AAY79209-23). The sequence was deduced from a cDNA clone (see  
 CC AA294211). Isolated from a gallbladder library. It shows homology to  
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is  
 CC expressed in dermatologic and gastrointestinal tissues,  
 CC especially those associated with inflammation and cell  
 CC proliferation. The new human transferases and polynucleotides can  
 CC be used in the diagnosis, prevention and treatment of cancer,  
 CC developmental disorders, gastrointestinal disorders, genetic  
 CC disorders, immunological disorders, neurological disorders,  
 CC reproductive disorders, and smooth muscle disorders. The  
 CC polypeptides can also be used to raise antibodies, and to screen

CC for agonists and antagonists of transferase activity.  
 XX Sequence 386 AA:  
 SQ Query Match 57.5%; Score 222; DB 21; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-214;  
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 165 EKACRSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVPRPRAVFRSRETKGDLMD 224  
 DB 165 EKACRSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVPRPRAVFRSRETKGDLMD 224  
 QY 225 SRVMOQHOKLKEKEDQPYVWQVYICQSOLEYTKTIQS.LPKALOERYLLVREDIARAPV 284  
 DB 225 SRVMOQHOKLKEKEDQPYVWQVYICQSOLEYTKTIQS.LPKALOERYLLVREDIARAPV 284  
 QY 285 ACTSNRYEFGLEFLPLDTWVHNITRGKMGDHAFTNARALNVNQAMRSLPYEYVS 344  
 DB 285 ACTSNRYEFGLEFLPLDTWVHNITRGKMGDHAFTNARALNVNQAMRSLPYEYVS 344  
 QY 345 RIQKACGDMNLGVRHVSSEQDRNLIDLITVTPEQIH 386  
 DB 345 RIQKACGDMNLGVRHVSSEQDRNLIDLITVTPEQIH 386  
 RESULT 5  
 AAB41947  
 ID AAB41947 standard; Protein: 418 AA.  
 XX AAB41947;  
 AC AAB41947;  
 XX 08-FEB-2001 (first entry)  
 DT Human CRFX CRF1711 polypeptide sequence SEQ ID NO:3422.  
 DE XX  
 XX Human: open reading frame; CRFX; detection; cytostatic; hepatotropic;  
 KW vulnerability; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiact;  
 KW immunostimulant; thrombolytic; coagulant; vasodilator; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antineuritic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200058473-AZ.  
 PN  
 XX  
 XX 05-OCT-2000.  
 PD  
 XX  
 XX 31-MAR-2000; 2000WO-US086621.  
 PE  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CUBA-) CUBAGEN CORP.  
 PA  
 XX  
 XX Shinkets RA, Leach M;  
 PI  
 XX  
 XX WPI: 2000-602362/57.  
 DR N-PSDB: AAC76156.  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2599-2600; 5507pp; English.  
 PS  
 XX AACT74446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human CRFX open reading frames 1 to 3161. The CRFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerability;  
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 CC anticonvulsant; osteopathic; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiact; thrombolytic; coagulant; vasodilator;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antineuritic; antithyroid; antineuritic;  
 CC antithyroid; antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an CRFX-associated disorder. The  
 CC nucleic acids can be used to express CRFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 SQ Sequence 418 AA;  
 Query Match 26.7%; Score 103; DB 21; Length 418;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-94;  
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 108 VFDAVMEGPPRROSLPQWENSRLCSAPACDIIPQDEIIPRAHCRCLCSQGFVEVERA 167  
 DB 134 VFDAVMEGPPRROSLPQWENSRLCSAPACDIIPQDEIIPRAHCRCLCSQGFVEVERA 193  
 QY 168 CRYSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVPRPRAV 210  
 DB 194 CRYSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVPRPRAV 236  
 RESULT 6  
 ABB81560  
 ID ABB81560 standard; Protein: 171 AA.  
 XX ABB81560;  
 AC ABB81560;  
 XX 05-SEP-2002 (first entry)  
 DT Human high endothelial cell GLCNAC6ST portion SEQ ID NO:8.  
 DE XX  
 XX Human: N-acetylglucosamine-6-sulfotransferase; enzyme; GLCNAC6ST;  
 KW corneal; sulfation; Keratan sulfate; macular corneal dystrophy; MCD;  
 KW ophthalmological.  
 KW  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX US2002061562-A1.  
 PN  
 XX  
 XX 23-MAY-2002.  
 PD  
 XX  
 XX 09-AUG-2001; 2001US-0927602.  
 PR 11-AUG-2000; 2000US-325773P.  
 PR  
 XX  
 XX (FUKU/) FUKUDA M N.  
 PA (AKAM/) AKAMA T O.  
 PA  
 XX  
 XX Fukuda MN, Akama TO;  
 PI  
 XX  
 XX WPI: 2002-507643/54.  
 DR  
 XX  
 XX New nucleic acid encoding corneal  
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment.

PT monitoring and diagnosis of macular corneal dystrophy -  
XX  
XX Example 2; Fig 3c; 69pp; English.  
PS  
XX The present invention describes human corneal  
CC N-acetylglicosamine-6-sulfotransferase (I), which is able to catalyze  
CC sulfation of keratan sulfate (KS). Also described is a method for  
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),  
CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,  
CC and has ophthalmological activity. (I) can be used to treat or prevent  
CC macular corneal dystrophy types I or II. (I) makes possible treatment  
CC of MCD without requiring keratinoplasty or keratectomy. The present  
CC sequence represents a portion of the human high endothelial cell  
CC N-acetylglicosamine-6-sulfotransferase (GlcNAc6ST), which is given in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 171 AA;  
Query Match 15.5%; Score 60; DB 23; Length 171;  
Best Local Similarity 100.0%; Pred. No. 9.5e-52;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 166 KACRSYSHVVLKVRFFNQSLYPLKDPSLNLIHVLYDRPRAVFRSRERTKGDLIMIDS 225  
DB 53 KACRSYSHVVLKVRFFNQSLYPLKDPSLNLIHVLYDRPRAVFRSRERTKGDLIMIDS 112  
RESULT 7  
AA39919  
ID AA39919 standard; Protein: 388 AA.  
XX  
AC AA39919;  
XX  
DT 08-DEC-1999 (first entry)  
XX  
DE Mouse glycosyl sulfotransferase-3 protein sequence.  
XX  
XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
KM selectin binding interaction; inflammation; lymphocyte homing; mouse;  
XX secondary lymph organ.  
XX  
OS Mus. sp.  
XX  
PN WO949018-A1.  
XX  
PD 30-SEP-1999.  
XX  
PF 26-FEB-1999; 99WO-US04316.  
XX  
PR 20-MAR-1998; 98US-0045284.  
PR 12-NOV-1998; 98US-0190911.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
PA (SYNT ) SYMTEX USA INC.  
XX  
PI Bistup A, Rosen SD, Tangemann K, Hemmerich S;  
XX  
XX WPI: 1999-580442/49.  
DR N-PSDB; AA220793.  
XX  
PT Human and murine glycosyl sulfotransferase 3 and related  
PT polynucleotides  
XX  
PS Claim 2; Fig 4; 59pp; English.  
XX  
XX This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of  
CC the invention. The nucleic acid sequences, probes and primers derived  
CC from these, proteins and antibodies are useful in detecting homologues.  
CC The sequences, antibodies and methods are useful in the diagnosis and  
CC treatment of diseases associated with selectin binding interactions,  
CC including conditions associated with or resulting from the homing of  
CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
CC to secondary lymph organs.

XX  
SQ Sequence 388 AA;  
Query Match 10.1%; Score 39; DB 20; Length 388;  
Best Local Similarity 100.0%; Pred. No. 2.3e-10;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 44 HVLVLSMRSGSFFVQGFQGHDPVFLMEPAMHWMTF 82  
DB 43 HVLVLSMRSGSFFVQGFQGHDPVFLMEPAMHWMTF 81  
RESULT 8  
ABB81558  
ID ABB81558 standard; Protein: 169 AA.  
XX  
AC ABB81558;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Human corneal N-acetylglicosamine-6-sulfotransferase portion SEQ.6.  
XX  
XX Human; N-acetylglicosamine-6-sulfotransferase; enzyme; GlcNAc6ST;  
KM corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
KM ophthalmological.  
XX  
OS Homo sapiens.  
XX  
PN US2002061562-A1.  
XX  
PD 23-MAY-2002.  
XX  
PF 09-AUG-2001; 2001US-0927602.  
XX  
PR 11-AUG-2000; 2000US-325773P.  
XX  
XX (FUKU/) FUKUDA M N.  
PA (AKAW/) AKAMA T O.  
XX  
PI Fukuda MN, Akama TO;  
XX  
DR WPI: 2002-507643/54.  
XX  
XX New nucleic acid encoding corneal  
PT N-acetylglicosamine-6-sulfotransferase, useful for treatment,  
PT monitoring and diagnosis of macular corneal dystrophy -  
XX  
XX Example 2; Fig 3c; 69pp; English.  
XX  
XX The present invention describes human corneal  
CC N-acetylglicosamine-6-sulfotransferase (II), which is able to catalyze  
CC sulfation of keratan sulfate (KS). Also described is a method for  
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),  
CC and detecting susceptibility to MCD. (II) is located to chromosome 16q22,  
CC and has ophthalmological activity. (II) can be used to treat or prevent  
CC macular corneal dystrophy types I or II. (II) makes possible treatment  
CC of MCD without requiring keratinoplasty or keratectomy. The present  
CC sequence represents a portion of the human corneal N-acetylglicosamine-  
CC 6-sulfotransferase (GlcNAc6ST), which is given in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 169 AA;  
Query Match 4.9%; Score 19; DB 23; Length 169;  
Best Local Similarity 100.0%; Pred. No. 1.2e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 44 HVLVLSMRSGSFFVQGF 62  
DB 10 HVLVLSMRSGSFFVQGF 28  
RESULT 9

ABB81559  
ID ABB81559 standard; Protein: 169 AA.  
XX  
AC ABB81559;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Human intestinal N-acetylglucosamine-6-sulfotransferase portion SEQ:7.  
XX  
KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GICNA6ST;  
KW Corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
KW Ophthalmological.  
XX  
OS Homo sapiens.  
XX  
PN US2002061562-A1.  
XX  
PD 23-MAY-2002.  
XX  
PF 09-AUG-2001; 2001US-0927602.  
XX  
PR 11-AUG-2000; 2000US-325773P.  
XX  
PA (EUKU/) FUKUDA M N.  
XX (AKAM/) AKAMA T O.  
XX  
PI Fukuda MN, Akama TO;  
XX  
DR WPI: 2002-507643/54.  
XX  
PT New nucleic acid encoding corneal  
PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,  
PT monitoring and diagnosis of macular corneal dystrophy  
XX  
XX Example 2; Fig 3C; 69pp; English.  
XX  
CC The present invention describes human corneal  
CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze  
CC sulfation of keratan sulfate (KS). Also described is a method for  
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),  
CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,  
CC and has ophthalmological activity. (I) can be used to treat or prevent  
CC macular corneal dystrophy types I or II. (I) makes possible treatment  
CC of MCD without requiring keratoplasty or keratectomy. The present  
CC sequence represents a portion of the human intestinal  
CC N-acetylglucosamine-6-sulfotransferase (GICNA6ST), which is given in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 169 AA:  
XX  
Query Match 4.9%; Score 19; DB 23; Length 169;  
Best Local Similarity 100.0%; Pred. No. 1,2e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 167 ACRSYSHVLEKEVRFNMQ 185  
DB 51 ACRSYSHVLEKEVRFNMQ 69  
RESULT 10  
AAV72639  
ID AAV72639 standard; Protein: 390 AA.  
XX  
AC AAV72639;  
XX  
DT 02-MAY-2001 (first entry)  
XX  
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha).  
XX  
KW Human: glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;  
KW therapy; selectin binding inhibitor; gene therapy; inflammation;  
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;  
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;  
KW

glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;  
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;  
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;  
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;  
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;  
KW chromosome 16q23.1.  
XX  
OS Homo sapiens.  
XX  
PN WO200106015-A1.  
XX  
PD 25-JAN-2001.  
XX  
PF 19-JUL-2000; 2000WO-US19741.  
XX  
PR 20-JUL-1999; 99US-0144694.  
XX 13-JUL-2000; 2000US-0593828.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Rosen SD, Lee JK, Hemmerich S;  
XX  
DR WPI: 2001-138471/14.  
XX N-PSDB: AAD02697, AAD02698, AAD02699.  
XX  
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for  
PT diagnostic and therapeutic agent screening applications  
XX  
XX Claim 3; Fig 1; 128pp; English.  
XX  
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4  
CC alpha). GST-4 gene is found on chromosome 16q23.1.  
CC GST is a type 2 membrane protein useful for inhibiting a binding event  
CC between selectin and a selectin ligand, which comprises contacting the  
CC selectin with a non-sulphated selectin ligand, GST and a small molecular  
CC agent that inhibits the sulphation activity of GST. GST is also useful  
CC in inhibiting a selectin mediated binding event. GST is useful in gene  
CC therapy to treat disorders such as acute or chronic inflammation,  
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis  
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's  
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious  
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,  
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
CC during transplantation.  
CC Note: The present sequence is also shown in sequence listing (page  
CC no: 56) but lacks four nucleotides at its 3' end.  
XX  
SQ Sequence 390 AA:  
XX  
Query Match 4.9%; Score 19; DB 22; Length 390;  
Best Local Similarity 100.0%; Pred. No. 2,6e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 167 ACRSYSHVLEKEVRFNMQ 185  
DB 165 ACRSYSHVLEKEVRFNMQ 183  
RESULT 11  
ABB81556  
ID ABB81556 standard; Protein: 390 AA.  
XX  
AC ABB81556;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.  
XX  
KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GICNA6ST;  
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
KW

PD 25-JAN-2001.  
 PE 19-JUL-2000; 2000WO-US19741.  
 PF 20-JUL-1999; 99US-0144694.  
 PR 13-JUL-2000; 2000US-0593828.  
 PS (REGC ) UNTV CALIFORNIA.  
 PT Rosen SD, Lee JK, Hemmerich S;  
 DR WPI: 2001-138471/14.  
 DR N-PSDB; AAD02696.  
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for  
 PT diagnostic and therapeutic agent screening applications  
 XX  
 XX  
 PS Claim 3; Fig 2; 128pp; English.  
 XX  
 XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4  
 CC gene is found on chromosome 8L1.  
 CC GST is a type 2 membrane protein useful for inhibiting a binding event  
 CC between a selectin and a selectin ligand, which comprises contacting the  
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular  
 CC agent that inhibits the sulphation activity of GST. GST is also useful  
 CC in inhibiting a selectin mediated binding event. GST is useful in gene  
 CC therapy to treat disorders such as acute or chronic inflammation,  
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis  
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's  
 CC disease, Grave's disease, adenallitis, hypoparathyroidism, pernicious  
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,  
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
 CC during transplantation.  
 CC  
 CC  
 SQ Sequence 395 AA:  
 XX  
 XX  
 XX Query Match 4.9%; Score 19; DB 22; Length 395;  
 XX Best Local Similarity 100.0%; Pred. No. 2,7e-10;  
 XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 44 HVLVLSWRSGSSEVQGLF 62  
 XX |||||||  
 DB 42 HVLVLSWRSGSSEVQGLF 60  
 XX  
 XX  
 XX RESULT 13  
 XX AAY72640  
 XX ID AAY72640 standard; Protein: 395 AA.  
 XX AAY72640;  
 XX  
 XX 02-MAY-2001 (first entry)  
 XX  
 XX Human glycosyl sulfotransferase-4beta (GST-4beta).  
 XX  
 XX Human: glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;  
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;  
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;  
 KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;  
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenallitis;  
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;  
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;  
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;  
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;  
 KW chromosome 16q23.1.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200106015-A1.  
 PN  
 XX

PD 25-JAN-2001.  
 XX  
 PF 19-JUL-2000; 2000WO-US19741.  
 XX  
 PR 20-JUL-1999; 99US-0144694.  
 PR 13-JUL-2000; 2000US-0593826.  
 XX  
 PA (RECC ) UNIV CALIFORNIA.  
 PI Rosen SD, Lee JK, Hemmerich S;  
 DR WPI: 2001-138471/14.  
 DR N-PSDB: AAD02697, AAD02700.  
 XX  
 PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for  
 PT diagnostic and therapeutic agent screening applications -  
 XX  
 PS Claim 3; Fig 4B; 128pp; English.  
 CC  
 CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4  
 CC beta). GST-4 gene is found on chromosome 16q23.1.  
 CC GST is a type 2 membrane protein useful for inhibiting a binding event  
 CC between a selectin and a selectin ligand, which comprises contacting the  
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular  
 CC agent that inhibits the sulphation activity of GST. GST is also useful  
 CC in inhibiting a selectin mediated binding event. GST is useful in gene  
 CC therapy to treat disorders such as acute or chronic inflammation,  
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis  
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's  
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious  
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,  
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
 CC during transplantation.  
 CC  
 XX  
 SQ Sequence 395 AA:  
 Query Match 4.9%; Score 19; DB 22; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 2,7e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 44 HVLVLSWRSGSSFFVGOLF 62  
 DB 42 HVLVLSWRSGSSFFVGOLF 60  
 RESULT 14  
 ABB81554  
 ID ABB81554 standard; Protein: 395 AA.  
 AC ABB81554;  
 XX  
 DT 05-SEP-2002 (first entry)  
 DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.  
 XX  
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNA6ST;  
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
 KW ophthalmological; chromosome 16q22.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002061562-A1.  
 XX  
 PD 23-MAY-2002.  
 XX  
 PF 09-AUG-2001; 2001US-0927602.  
 XX  
 PR 11-AUG-2000; 2000US-325773P.  
 XX  
 PA (FUKU/) FUKUDA M N.

PA (AKAM/) AKAMA T O.  
 XX  
 PI Fukuda MN, Akama TO;  
 XX  
 DR WPI: 2002-507643/54.  
 DR N-PSDB: ABN89506.  
 XX  
 PT New nucleic acid encoding corneal  
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,  
 PT monitoring and diagnosis of macular corneal dystrophy -  
 XX  
 PS Claim 13; Fig 1A-D; 69pp; English.  
 CC  
 CC The present sequence represents human corneal  
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze  
 CC sulfation of keratan sulfate (KS). Also described is a method for  
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),  
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,  
 CC and has ophthalmological activity. (I) can be used to treat or prevent  
 CC macular corneal dystrophy types I or II. (I) makes possible treatment  
 CC of MCD without requiring keratoplasty or keratectomy.  
 CC  
 XX  
 SQ Sequence 395 AA:  
 Query Match 4.9%; Score 19; DB 23; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 2,7e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 44 HVLVLSWRSGSSFFVGOLF 62  
 DB 42 HVLVLSWRSGSSFFVGOLF 60  
 RESULT 15  
 ABB81555  
 ID ABB81555 standard; Protein: 395 AA.  
 AC ABB81555;  
 XX  
 DT 05-SEP-2002 (first entry)  
 DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.  
 XX  
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNA6ST;  
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
 KW ophthalmological.  
 XX  
 OS Homo sapiens.  
 OS Mus musculus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT MISC-difference 10 /label= Ala, Thr, Val  
 FT MISC-difference 13 /label= Ala, Val, Ser  
 FT MISC-difference 20 /label= Phe, Cys, Gly  
 FT MISC-difference 39 /label= Ala, Asp, Glu  
 FT MISC-difference 96 /label= Val, Met, Ile  
 FT MISC-difference 142 /label= Ala, Thr, Asn  
 FT MISC-difference 147 /label= Ala, Asp, Glu  
 FT MISC-difference 159 /label= Thr, Ser, Gly  
 FT MISC-difference 238 /label= Gly, His, Arg  
 FT MISC-difference 294 /label= Ser, Thr, Gly  
 FT MISC-difference 371



FT Misc-difference /label= Ala, Thr, Ser  
 FT Misc-difference 380 /label= Leu, Pro, Met  
 FT Misc-difference 382 /label= Gly, His, Ser  
 FT Misc-difference 384 /label= Thr, Ser, Lys  
 FT Misc-difference 390 /label= Ala, Glu  
 FT Misc-difference 391 /label= Ser, Lys  
 FT Misc-difference 392 /label= His, Gln  
 FT Misc-difference 394 /label= Arg, Glu  
 FT Misc-difference 395 /label= Asn, Ser  
 FT US2002061562-A1.  
 PD 23-MAY-2002.  
 PF 09-AUG-2001; 2001US-0927602.  
 PR 11-AUG-2000; 2000US-325773P.  
 PA (FUKU/) FUKUDA M N.  
 PA (AKAM/) AKAMA T O.  
 XX Fukuda MN, Akama TO;  
 PI WPI: 2002-507643/54.  
 DR  
 XX  
 PT New nucleic acid encoding corneal  
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,  
 PT monitoring and diagnosis of macular corneal dystrophy -  
 XX  
 PS Example 5; Fig 2A-B; 69pp; English.  
 XX  
 CC The present invention describes human corneal  
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse  
 CC sulfation of keratan sulfate (KS). Also described is a method for  
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),  
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,  
 CC and has ophthalmological activity. (I) can be used to treat or prevent  
 CC macular corneal dystrophy types I or II. (I) makes possible treatment  
 CC of MCD without requiring keratoplasty or keratectomy. The present  
 CC sequence represents a consensus N-acetylglucosamine-6-sulfotransferase  
 CC which is given in the exemplification of the present invention.  
 XX  
 SO Sequence 395 AA;  
 Query Match 4.9%; Score 19; DB 23; Length 395;  
 Best Locality 100.0%; Pred. No. 2.7e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 44 HVLVLSMRSGSSFGOLF 62  
 Db 42 HVLVLSMRSGSSFGOLF 60  
 RESULT 16  
 AAE15438  
 ID AAE15438 standard; Protein; 395 AA.  
 XX  
 AC AAE15438;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Human drug metabolizing enzyme (DME)-5.  
 XX  
 KW Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;  
 KW Inflammatory disorder; acquired immune deficiency syndrome; infection;

KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;  
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;  
 KW cancer; endocrine disorder; hypochalams disorder; pituitary disorder;  
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;  
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;  
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;  
 KW DME-5.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..32  
 FT /label= Signal\_peptide  
 FT Protein 33..395  
 FT /note= "Human mature DME-5 protein"  
 FT Peptide 1..35  
 FT /label= Signal\_peptide  
 FT Protein 36..395  
 FT /note= "Human mature DME-5 protein"  
 XX  
 XX WO200179468-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 12-APR-2001; 2001WO-0511869.  
 XX  
 PR 13-APR-2000; 2000US-197590P.  
 PR 19-APR-2000; 2000US-198403P.  
 PR 28-APR-2000; 2000US-200185P.  
 PR 05-MAY-2000; 2000US-202234P.  
 PR 11-MAY-2000; 2000US-203509P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Policky JL, Hatalla A, Burford N, Ring HZ, Lal P, Tribouley CM;  
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;  
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;  
 PI Au-Young J;  
 XX  
 DR WPI: 2002-066363/09.  
 DR N-PSDB: AAD24670.  
 XX  
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,  
 PT useful for diagnosing, treating, or preventing disorders associated  
 PT with aberrant expression of DME such as allergy, anemia, asthma,  
 PT infertility -  
 XX  
 PS Claim 1a; Page 131-132; 143pp; English.  
 XX  
 CC The invention relates to human drug metabolizing enzymes referred as  
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides  
 CC of the invention are useful for assessing toxicity of test compounds  
 CC and in gene therapy. Sequences of the invention are useful in the  
 CC diagnosis, prevention and treatment of autoimmune/inflammatory  
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult  
 CC respiratory distress syndrome, allergies, anemia, atherosclerosis,  
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's  
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's  
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple  
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus  
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,  
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic  
 CC infections; cell proliferative disorders such as actinic keratosis,  
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,  
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,  
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus  
 CC and pituitary resulting from lesions such as primary brain tumours,  
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular  
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,  
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,  
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,  
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,  
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such

CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,  
CC hyperbilirubinemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's  
CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental  
CC disorders. The present sequence is human DMF-5 protein.

XX Sequence 395 AA;

Query Match 4.9%; Score 19; DB 23; Length 395;  
Best Local Similarity 100.0%; Pred. No. 2,7e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRSGSSFFVGOLF 62

DB 42 HVLVSSWRSGSSFFVGOLF 60

RESULT 17

AAU11275

ID AAU11275 standard; Protein; 395 AA.

XX AAU11275;

DT 12-MAR-2002 (first entry)

DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX Mouse; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;  
XX L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;  
XX ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;  
XX allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;  
XX delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;  
XX antiinflammatory; antipsoriatic; antidiabetic; dermatological;  
XX antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.

XX Mus musculus.

XX WO200185177-A1.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001WO-US15452.

XX 11-MAY-2000; 2000US-0569320.

XX (BURN-) BURHAM INST.

XX Fukuda M, Yeh J, Hirooka N;

XX WPI; 2002-075226/10.

XX N-PSDB; AAS16948.

PT New enzyme, useful for modifying acceptor molecule, comprises an  
PT isolated L-selectin sulfotransferase-2 that directs expression of  
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or  
PT Intestinal GlcNAc 6-sulfotransferase

XX Claim 28; Fig 10; 98pp; English.

XX The present invention provides a method of modifying an acceptor molecule  
XX by contacting the acceptor with an isolated  
XX beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active  
XX fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The  
XX invention also provides a method of treating or preventing an  
XX L-selectin-mediated condition by reducing the expression or activity of a  
XX beta1,3GNT that directs expression of a MECA-79 antigen. This can be done  
XX by administering to the subject an oligosaccharide L-selectin antagonist  
XX that inhibits binding of L-selectin to a MECA-79 antigen, for example by  
XX administering antibody material that specifically binds beta1,3GNT,  
XX and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin  
XX sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.  
XX Alternatively, the expression or activity of LST-2 or its active  
XX fragment can be reduced in combination with reducing the expression or  
XX activity of beta1,3GNT. The method is useful for treating L-selectin

CC mediated conditions such as Crohn's disease and ulcerative colitis,  
CC inflammatory disorders of the skin such as allergic contact dermatitis,  
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
CC sequence represents mouse I-GlcNAc6ST.

XX Sequence 395 AA;

Query Match 4.9%; Score 19; DB 23; Length 395;  
Best Local Similarity 100.0%; Pred. No. 2,7e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRSGSSFFVGOLF 62

DB 42 HVLVSSWRSGSSFFVGOLF 60

RESULT 18

ABB81557

ID ABB81557 standard; Protein; 418 AA.

XX ABB81557;

DT 05-SEP-2002 (first entry)

DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.

XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;  
XX Korneel; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
XX ophthalmological.

XX Mus musculus.

XX US2002061562-A1.

XX 23-MAY-2002.

XX 09-AUG-2001; 2001US-0927602.

XX 11-AUG-2000; 2000US-325773P.

XX (FUKU/) FUKUDA M N.

XX (AKAM/) AKAMA T O.

XX Fukuda MN, Akama TO;

XX WPI; 2002-507643/54.

PT New nucleic acid encoding corneal  
PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,  
PT monitoring and diagnosis of macular corneal dystrophy  
XX Example 5; Page 24-25; 69pp; English.

XX The present invention describes human corneal  
XX N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze  
XX sulfation of keratan sulfate (KS). Also described is a method for  
XX monitoring the effect of treatments for macular corneal dystrophy (MCD),  
XX and detecting susceptibility to MCD. (I) is located to chromosome 16q22,  
XX and has ophthalmological activity. (I) can be used to treat or prevent  
XX macular corneal dystrophy types I or II. (I) makes possible treatment  
XX of MCD without requiring keratoplasty or keratotomy. The present  
XX sequence represents mouse intestinal N-acetylglucosamine-6-  
XX sulfotransferase, which is given in comparison with (I) in the  
XX exemplification of the present invention.

XX Sequence 418 AA;

Query Match 4.9%; Score 19; DB 23; Length 418;  
Best Local Similarity 100.0%; Pred. No. 2,8e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRSGSSFFVGOLF 62

Mon Jan 13 09:14:02 2003

us-09-816-825-2.0115.rag

Page 11

Db 66 HVLVLSWRSSSFVGOLF 84

Search completed: January 11, 2003, 01:29:48  
Job time : 37 secs

**This Page Blank (uspto)**

Mon Jan 13 09:14:04 2003

us-09-816-825-2.olil5.rpr

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:29:11 : Search time 19 seconds  
(without alignments)  
1953.047 Million cell updates/sec

Title: US-09-816-825-2  
Perfect score: 386  
Sequence: 1 MULPKKKMLLFLVSQMAIL.....EGRNLLDLLSTWVPEQIH 386

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	ID	Description
No matches found				

Search completed: January 11, 2003, 01:31:32  
Job time : 19 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:23:06 ; Search time 13 Seconds  
(without alignments)  
1231.528 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 386  
Sequence: 1 MLIPKMKLLFLVSOMAIL.....EQRNLLDLSTWVPEQIH 386

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query				
No.	Score	Match	length	ID	Description
-----					
No matches found					

Search completed: January 11, 2003, 01:30:15  
Job time : 13 secs

**This Page Blank (uspto)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:23:56 ; Search time 31 seconds

(without alignments)  
2565.619 Million cell updates/sec

Title: US-09-816-825-2

Sequence: 1 MLPPKKMLLFVSQMAIL.....EQRNLLDLSTWVPEQIH 386

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size: 15

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

SPREMBL\_21:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	386	100.0	386	4	Q9Y5R3
2	39	10.1	388	11	Q9R1I1
3	39	10.1	388	11	Q9WDE5
4	19	4.9	395	4	Q9GZX3
5	19	4.9	395	11	Q9QUP4
6	19	4.9	411	4	Q9GZS9

#### ALIGNMENTS

RESULT 1  
ID Q9Y5R3 PRELIMINARY: PRT: 386 AA.  
AC Q9Y5R3;  
DT 01-NOV-1999 (TRENBLREL, 12, Created)

DT 01-NOV-1999 (TRENBLREL, 12, Last sequence update)  
01-MAR-2002 (TRENBLREL, 20, Last annotation update)  
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand  
DE sulfotransferase GST-3).  
GN GST3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TONSIL;  
RX MEDLINE=99264336; PubMed=10330415;  
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,  
Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;  
RT "Sulfotransferases of two specificities function in the reconstitution  
of high endothelial cell ligands for L-selectin."  
J. Cell Biol. 145:899-910(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TONSIL;  
RX MEDLINE=21332592; PubMed=11439191;  
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Elies L.G.,  
Rabuka D., Hindsaul O., Marth J.D., Lowe J.B., Fakuda M.;  
RT "Novel sulfated lymphocyte homing receptors and their control by a  
corel extension beta1,3-N-acetylglucosaminyltransferase."  
Cell 105:957-969(2001).  
RL EMBL: AF131235; AAC33015.1; -  
DR EMBL: AF280088; AAG48246.1; -  
DR EMBL: AF149783; AAK48417.1; -  
DR InterPro: IPR001092; HLH\_basic.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
KW Lectin; Selectin; Transferase.  
SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417743A CRC64;

Query Match 100.0%; Score 386; DB 4; Length 386;

Best Local Similarity 100.0%; Pred. NO. 0;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLPPKKMLLFVSQMAILFFHNYSHNSSLKKAQPERHAYVLSMRSGSFFVQ	60
DB	1	MLPPKKMLLFVSQMAILALFFHNYSHNSSLKKAQPERHAYVLSMRSGSFFVQ	60
QY	61	LFQGHDPVEYLMEPAHWMTFKQSTAWMLHMAVDLIRAVFLCDMSVDFAYMEPGRQ	120
DB	61	LFQGHDPVEYLMEPAHWMTFKQSTAWMLHMAVDLIRAVFLCDMSVDFAYMEPGRQ	120
QY	121	SSLFQWENSRLCSAPACDIIPODEIIPRAHGRLLCSQPFVEVEKACSYSVYLKEVR	180
DB	121	SSLFQWENSRLCSAPACDIIPODEIIPRAHGRLLCSQPFVEVEKACSYSVYLKEVR	180
QY	181	FFNIQSLYPLKDPSSLNLHVLVDRPRAVFRSREKTKGDLMSRIYVQGHQKLKED	240
DB	181	FFNIQSLYPLKDPSSLNLHVLVDRPRAVFRSREKTKGDLMSRIYVQGHQKLKED	240
QY	241	QPYVVMQVTCQSOLEIYKTQSIPKALQERYLLVREDLARAFAQTSRMYEFGLEFLP	300
DB	241	QPYVVMQVTCQSOLEIYKTQSIPKALQERYLLVREDLARAFAQTSRMYEFGLEFLP	300
QY	301	HLQTVWNITRGKGMGDAFTNARDALNYSQAMPWSLPEYKYSRLQKACGDAAMLDCYR	360
DB	301	HLQTVWNITRGKGMGDAFTNARDALNYSQAMPWSLPEYKYSRLQKACGDAAMLDCYR	360
QY	361	HWSEQEQRNLLDLSTWVPEQIH 386	

Db 361 HVHSEGEORNLDDLSTWTPVPEQIH 386

# RESULT 2

AC 09R111; PRELIMINARY; PRT; 388 AA.  
 AC 09R111; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE L-selectin ligand sulfotransferase.  
 GN CHST4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99361934; PubMed=10435581;  
 RA Hirooka N., Petrnyak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,  
 RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;  
 RT A novel, high endothelial venule-specific sulfotransferase expresses  
 RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";  
 RL Immunity 11:79-89(1999).  
 DR EMBL: AF109155; Add4579.1;  
 DR MGD: MGI:1349479; Chst4.  
 KM Lectin; Selectin; Transferase.  
 SQ SEQUENCE 388 AA; 44636 MW; 6D5371AFB684AEE CRC64;

Query Match 10.1%; Score 39; DB 11; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-34;  
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 HVLVLSWRSGSFVGLFGQHPDVFYLMEPANVMWTF 82  
 Db 43 HVLVLSWRSGSFVGLFGQHPDVFYLMEPANVMWTF 81

# RESULT 3

AC 09WDE5; PRELIMINARY; PRT; 388 AA.  
 AC 09WDE5; 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (chondroitin  
 DE 6-Keratan) sulfotransferase 4).  
 GN CHST4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6; TISSUE=TONSIL;  
 RX MEDLINE=99264336; PubMed=10330415;  
 RA Bistup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,  
 RA Huang C.-C., Kanagaki R., Rosen S.D., Hemmerich S.;  
 RT Sulfotransferases of two specificities function in the reconstitution  
 RT of high endothelial cell ligands for L-selectin.";  
 RL J. Cell Biol. 145:899-910(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuenl P., Lewis S., Matsuo Y., Nakano I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bona M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Holman M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszewski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AF131236; Add33016.1;  
 DR EMBL: AK009113; BAB26078.1;  
 DR MGD: MGI:1349479; Chst4.  
 KV Transferase.  
 SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 10.1%; Score 39; DB 11; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-34;  
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 HVLVLSWRSGSFVGLFGQHPDVFYLMEPANVMWTF 82  
 Db 43 HVLVLSWRSGSFVGLFGQHPDVFYLMEPANVMWTF 81

# RESULT 4

AC 09GZX3; PRELIMINARY; PRT; 395 AA.  
 AC 09GZX3; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-  
 DE acetylglucosamine-6-O-sulfotransferase).  
 GN GST4BETA OR CHST6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistup A., Ruddle N.R.,  
 RA Rosen S.D.;  
 RT "Chromosomal Localization and Genomic Organization for the  
 RT Galactose/N-acetylgalactosamine/N-Acetylglucosamine 6-O-  
 RT Sulfotransferase Gene Family.";  
 RL Glycobiology 0:0-0(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20472330; PubMed=11017086;  
 RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,  
 RA Nakamura T., Doi A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,  
 RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;  
 RT "Mucular corneal dystrophy type I and type II are caused by distinct  
 RT mutations in a new sulphotransferase gene.";  
 RL Nat. Genet. 26:237-241(2000).  
 DR EMBL: AF280086; AAC4824.1;  
 DR EMBL: AF219980; AAC26325.1;  
 DR EMBL: AF219991; AAC26327.1;  
 KV Transferase.  
 SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 4.9%; Score 19; DB 4; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 HVLVLSWRSGSFVGLFGQHPDVFYLMEPANVMWTF 82  
 Db 42 HVLVLSWRSGSFVGLFGQHPDVFYLMEPANVMWTF 60

RESULT 5  
 ID 090UP4 PRELIMINARY: PRT: 395 AA.  
 AC 090UP4:  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE N-acetylglucosamine 6-O-sulfotransferase.  
 GN CHST5 OR I-GLCNAC-6-ST.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=INTESTINE;  
 RX MEDLINE=99423499; PubMed=10491328;  
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;  
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-  
 Sulfotransferase that is Highly Restricted to Intestinal Tissue.";  
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).  
 DR EMBL/AF176841; AAD56003.1; -  
 DR EMBL/AF176840; AAD56002.1; -  
 DR MGD:MGI:1931825; Chst5.  
 KW Transferase.  
 SQ SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;

Query Match 4.9%; Score 19; DB 11; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 HVLVSSMRSGSFFVQGF 62  
 |||  
 DB 42 HVLVSSMRSGSFFVQGF 60

## RESULT 6

ID 09GZS9 PRELIMINARY: PRT: 411 AA.  
 AC 09GZS9; 09UBY3;

DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Intestinal GLCNAC-6-sulfotransferase (Intestinal N-  
 acetylglucosamine-6-O-sulfotransferase) (N-acetylglucosamine  
 6-O-sulfotransferase).  
 GN CHST5 OR I-GLCNAC-6-ST.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20472350; PubMed=11017086;  
 RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,  
 RA Nakamura T., Dola A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,  
 RA Fujiwara T., Rhonara E.J., Shimomura Y., Kinoshita S., Tanigami A.,  
 RA Fukuda M.N.;  
 RT "Macular corneal dystrophy type I and type II are caused by distinct  
 mutations in a new sulphotransferase gene.";  
 RL Nat. Genet. 26:237-241(2000).  
 RN [2]

RP SEQUENCE OF 22-411 FROM N.A.  
 RC TISSUE=INTESTINE;  
 RX MEDLINE=99423499; PubMed=10491328;  
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;  
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-  
 Sulfotransferase that is Highly Restricted to Intestinal Tissue.";  
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).  
 DR EMBL/AF246718; AAG28023.1; -  
 DR EMBL/AF19991; AAG26326.1; -  
 DR EMBL/AF176839; AAD56001.1; -  
 DR EMBL/AF176838; AAD56000.1; -

KW Transferase.  
 SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;  
 Query Match 4.9%; Score 19; DB 4; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 ACRSYSHVYKREFFNLQ 185  
 |||  
 DB 186 ACRSYSHVYKREFFNLQ 204

Search completed: January 11, 2003, 01:31:00  
 Job time : 32 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:29:56 ; Search time 16 Seconds

(without alignments)  
709,828 Million cell updates/sec

Title: US-09-816-825-2

Sequence: 1 MLPPKKKLLFLVSQMAIL.....EQNLDDLSTWTFEQIH 386

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size: 15

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	386	100.0	386	4	US-09-045-284A-2
2	386	100.0	386	4	US-09-190-911-1

#### ALIGNMENTS

RESULT 1  
US-09-045-284A-2  
; Sequence 2, Application US/09045284A  
; Patent No. 6265192  
; GENERAL INFORMATION:  
; APPLICANT: Bistrup, Annette  
; APPLICANT: Rosen, Steven D.  
; APPLICANT: Hemmerich, Stefan  
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
; FILE REFERENCE: 6510-107051  
; CURRENT APPLICATION NUMBER: US/09/045,284A  
; CURRENT FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 386  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-045-284A-2

Query Match 100.0%; Score 386; DB 4; Length 386;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLPPKKKLLFLVSQMAILALFFHMYSHNSSLKKAQPERMHVYLSSMRGSSSFVQ	60
DB	1	MLPPKKKLLFLVSQMAILALFFHMYSHNSSLKKAQPERMHVYLSSMRGSSSFVQ	60
QY	61	LFQGHDPVFLYLPBAMVWMTFKOSTAMLMHAYRDLIRAVFLCDMSVDPAYMEPPRRQ	120
DB	61	LFQGHDPVFLYLPBAMVWMTFKOSTAMLMHAYRDLIRAVFLCDMSVDPAYMEPPRRQ	120
QY	121	SSLFQWENSRALCSAPACDIIPDEIIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEVR	180
DB	121	SSLFQWENSRALCSAPACDIIPDEIIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEVR	180
QY	181	FFNLQSLYPLKDPSTLHIVHVRDPAVFRSRERTKGLMDSIYMGQHEQKLKED	240
DB	181	FFNLQSLYPLKDPSTLHIVHVRDPAVFRSRERTKGLMDSIYMGQHEQKLKED	240
QY	241	QPYVWGVICQSQLEIKTKIQSLPKAQERYLLVREDLARAFAVADTSRYEFVGLFLP	300
DB	241	QPYVWGVICQSQLEIKTKIQSLPKAQERYLLVREDLARAFAVADTSRYEFVGLFLP	300
QY	301	HLQTVHNITRGKMGDAFHTNARDALNYSQAMRSLPEYKYSRLQKACGDAMNLLGYR	360
DB	301	HLQTVHNITRGKMGDAFHTNARDALNYSQAMRSLPEYKYSRLQKACGDAMNLLGYR	360
QY	361	HVRSEQQRNLDDLSTWTFEQIH 386	
DB	361	HVRSEQQRNLDDLSTWTFEQIH 386	

RESULT 2  
US-09-190-911-1  
; Sequence 1, Application US/09190911  
; Patent No. 6365365  
; GENERAL INFORMATION:  
; APPLICANT: Bistrup, Annette  
; APPLICANT: Rosen, Steven D.  
; APPLICANT: Tangemann, Kirsten  
; APPLICANT: Hemmerich, Stefan  
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
; FILE REFERENCE: 6510-107CIP  
; CURRENT APPLICATION NUMBER: US/09/190,911  
; CURRENT FILING DATE: 1998-11-12  
; EARLIER APPLICATION NUMBER: 09/045,284  
; EARLIER FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 386  
; TYPE: PRT  
; ORGANISM: H. sapiens  
US-09-190-911-1

Query Match 100.0%; Score 386; DB 4; Length 386;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLPPKKKLLFLVSQMAILALFFHMYSHNSSLKKAQPERMHVYLSSMRGSSSFVQ	60
DB	1	MLPPKKKLLFLVSQMAILALFFHMYSHNSSLKKAQPERMHVYLSSMRGSSSFVQ	60
QY	61	LFQGHDPVFLYLPBAMVWMTFKOSTAMLMHAYRDLIRAVFLCDMSVDPAYMEPPRRQ	120
DB	61	LFQGHDPVFLYLPBAMVWMTFKOSTAMLMHAYRDLIRAVFLCDMSVDPAYMEPPRRQ	120
QY	121	SSLFQWENSRALCSAPACDIIPDEIIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEVR	180
DB	121	SSLFQWENSRALCSAPACDIIPDEIIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEVR	180
QY	181	FFNLQSLYPLKDPSTLHIVHVRDPAVFRSRERTKGLMDSIYMGQHEQKLKED	240

Db 181 FENLOSXPPLKDPSSLNLHIVLVDPRVFRSRERITKGDLMIDSRIMGQHEQKKED 240  
QY 241 QPYVMQVYCOSQLEITYKTISLPRALQERYLIVREDLARAPVAQTSRMVEFVGLLELP 300  
Db 241 QPYVMQVYCOSQLEITYKTISLPRALQERYLIVREDLARAPVAQTSRMVEFVGLLELP 300  
QY 301 HLOTWVHNITRGKMGDHAFTNARDALNVSQAWRWSLPEKVSRLQKACGDAMNLLGYR 360  
Db 301 HLOTWVHNITRGKMGDHAFTNARDALNVSQAWRWSLPEKVSRLQKACGDAMNLLGYR 360  
QY 361 HVRSQEQORNLDDLSTWTVEQIH 386  
Db 361 HVRSQEQORNLDDLSTWTVEQIH 386

Search completed: January 11, 2003, 01:32:03  
Job time : 17 secs

Mon Jan 13 09:14:03 2003

us-09-816-825-2.cl115.rapb

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:30:21 Search time 13 Seconds  
(without alignments)  
576.062 Million cell updates/sec

Title: US-09-816-825-2  
Perfect score: 386  
Sequence: 1 MLPPKMKLLFLVSQMAIL.....EQRNLLDLLSTWTPEDQH 386

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 118974 segs, 19401057 residues

Word size: 15

Total number of hits satisfying chosen parameters: 9

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

Published Applications-AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	386	100.0	386	9	US-10-007-262-1
2	386	100.0	386	10	US-09-816-825-2
3	60	15.5	171	10	US-09-927-602-8
4	19	4.9	169	10	US-09-927-602-6
5	19	4.9	169	10	US-09-927-602-7
6	19	4.9	390	10	US-09-927-602-4
7	19	4.9	395	10	US-09-927-602-2
8	19	4.9	395	10	US-09-927-602-3
9	19	4.9	418	10	US-09-927-602-5

# ALIGNMENTS

RESULT 1  
US-10-007-262-1  
Sequence 1, Application US/10007262  
Patent No. US20020164748A1  
GENERAL INFORMATION:

APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Tangemann, Kirsten  
APPLICANT: Hemmerich, Stefan  
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107CIP  
CURRENT APPLICATION NUMBER: US/10/007,262  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12  
NUMBER OF SEQ. ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ. ID NO. 1  
LENGTH: 386  
TYPE: PRT  
ORGANISM: H. sapiens  
US-10-007-262-1

Query Match 100.0%; Score 386; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSQMAILALFFHMYSHNSSLKKAQPERHVLVLSWRGSSFFVGQ 60  
DB 1 MLPPKMKLLFLVSQMAILALFFHMYSHNSSLKKAQPERHVLVLSWRGSSFFVGQ 60  
QY 61 LFGQHPDVEFLMEPAHWMTTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRQ 120  
DB 61 LFGQHPDVEFLMEPAHWMTTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRQ 120  
QY 121 SLPFQWNSALCSAPACDIPDEIIPRAHCHLLCSQPFVEVERACSSYHVLKEFR 180  
DB 121 SLPFQWNSALCSAPACDIPDEIIPRAHCHLLCSQPFVEVERACSSYHVLKEFR 180  
QY 181 FPNQSLYPLKPSLNLHVLVPRPRAVRSRERTKGDMLDSRIVGQHEOKLKED 240  
DB 181 FPNQSLYPLKPSLNLHVLVPRPRAVRSRERTKGDMLDSRIVGQHEOKLKED 240  
QY 241 QPYVWQVTOQSLEIYKTIQSLPKALQERYLLVRREDLARAFAVQTSRYEFGLEFLP 300  
DB 241 QPYVWQVTOQSLEIYKTIQSLPKALQERYLLVRREDLARAFAVQTSRYEFGLEFLP 300  
QY 301 HLOTWVNIITRGKMGDPHETNARDALNVSQAMRSLPEYKYSRLQKAGDAMNLGYR 360  
DB 301 HLOTWVNIITRGKMGDPHETNARDALNVSQAMRSLPEYKYSRLQKAGDAMNLGYR 360  
QY 361 HVRSQEQQRNLLDLLSTWTPEDQH 386  
DB 361 HVRSQEQQRNLLDLLSTWTPEDQH 386

RESULT 2  
US-09-816-825-2  
Sequence 2, Application US/09816825  
Patent No. US20010051370A1  
GENERAL INFORMATION:  
APPLICANT: Bistrup, Annette  
APPLICANT: Rosen, Steven D.  
APPLICANT: Hemmerich, Stefan  
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
FILE REFERENCE: 6510-107CON  
CURRENT APPLICATION NUMBER: US/09/816,825  
CURRENT FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/045,284  
PRIOR FILING DATE: 1998-03-20  
NUMBER OF SEQ. ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ. ID NO. 2  
LENGTH: 386  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-816-825-2

Query Match 100.0%; Score 386; DB 10; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPRKMLLFVLSQMAILALFFHMSHNTSSLSKRAQPRMHLVLSWSSGSSSVGQ 60  
 DB 1 MLPRKMLLFVLSQMAILALFFHMSHNTSSLSKRAQPRMHLVLSWSSGSSSVGQ 60  
 QY 61 LFGQPDVFLYMEPRMHWMTFKOSTAMLMHAYRDLIRAVFLCDMSYDPAVMPGRRQ 120  
 DB 61 LFGQPDVFLYMEPRMHWMTFKOSTAMLMHAYRDLIRAVFLCDMSYDPAVMPGRRQ 120  
 QY 121 SLEFGWNSRALCSAPACDIIPODEIIPRAHCRLLCSQOPPEVEVEKACRSYSHVYLKEVR 180  
 DB 121 SLEFGWNSRALCSAPACDIIPODEIIPRAHCRLLCSQOPPEVEVEKACRSYSHVYLKEVR 180  
 QY 181 FENQSLTEPLKDPDLNLHIHLVDRPRAVRSRRTKGDMLDSRYMGHBEKLEKED 240  
 DB 181 FENQSLTEPLKDPDLNLHIHLVDRPRAVRSRRTKGDMLDSRYMGHBEKLEKED 240  
 QY 241 QPYVYMOYICQSOLEIKTTIQSLPRALQERYLLVRYEDLARAPYAQTSRMVEFYGLEFLP 300  
 DB 241 QPYVYMOYICQSOLEIKTTIQSLPRALQERYLLVRYEDLARAPYAQTSRMVEFYGLEFLP 300  
 QY 301 HLOQWVHNTKRGKMGDAFTNARDALNVSQAMRSLPYEKYSRLKACGDANLLGYR 360  
 DB 301 HLOQWVHNTKRGKMGDAFTNARDALNVSQAMRSLPYEKYSRLKACGDANLLGYR 360  
 QY 361 HVRSQEQRNLLDLSTWTWPEQIH 386  
 DB 361 HVRSQEQRNLLDLSTWTWPEQIH 386

RESULT 3  
 US-09-927-602-8  
 ; Sequence 8, Application US/09927602  
 ; Patent No. US20020061562A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fukuda, Michiko N.  
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal  
 ; FILE REFERENCE: P-LJ 4852  
 ; CURRENT APPLICATION NUMBER: US/09/927,602  
 ; CURRENT FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: US 09/638,211  
 ; PRIOR FILING DATE: 2000-08-11  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 8  
 ; LENGTH: 171  
 ; TYPE: PRF  
 ; ORGANISM: Homo Sapien  
 US-09-927-602-8

Query Match 15.5%; Score 60; DB 10; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-50;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KACRSYSHVYLKEVRFNLSQSLPLKDPDLNLHIHLVDRPRAVRSRRTKGDMLDS 225  
 DB 53 KACRSYSHVYLKEVRFNLSQSLPLKDPDLNLHIHLVDRPRAVRSRRTKGDMLDS 112  
 RESULT 4  
 US-09-927-602-6  
 ; Sequence 6, Application US/09927602  
 ; Patent No. US20020061562A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fukuda, Michiko N.  
 ; APPLICANT: Akama, Tomoya O.  
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal  
 ; TITLE OF INVENTION: Dystrophy

FILE REFERENCE: P-LJ 4852  
 ; CURRENT APPLICATION NUMBER: US/09/927,602  
 ; CURRENT FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: US 09/638,211  
 ; PRIOR FILING DATE: 2000-08-11  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6  
 ; LENGTH: 169  
 ; TYPE: PRF  
 ; ORGANISM: Homo Sapien  
 US-09-927-602-6

Query Match 4.9%; Score 19; DB 10; Length 169;  
 Best Local Similarity 100.0%; Pred. No. 8e-11;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSWSSGSSFGOLF 62  
 DB 10 HVLVLSWSSGSSFGOLF 28

RESULT 5  
 US-09-927-602-7  
 ; Sequence 7, Application US/09927602  
 ; Patent No. US20020061562A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fukuda, Michiko N.  
 ; APPLICANT: Akama, Tomoya O.  
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal  
 ; FILE REFERENCE: P-LJ 4852  
 ; CURRENT APPLICATION NUMBER: US/09/927,602  
 ; CURRENT FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: US 09/638,211  
 ; PRIOR FILING DATE: 2000-08-11  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 169  
 ; TYPE: PRF  
 ; ORGANISM: Homo Sapien  
 US-09-927-602-7

Query Match 4.9%; Score 19; DB 10; Length 169;  
 Best Local Similarity 100.0%; Pred. No. 8e-11;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ACRSYSHVYLKEVRFNLSQ 185  
 DB 51 ACRSYSHVYLKEVRFNLSQ 69

RESULT 6  
 US-09-927-602-4  
 ; Sequence 4, Application US/09927602  
 ; Patent No. US20020061562A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fukuda, Michiko N.  
 ; APPLICANT: Akama, Tomoya O.  
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal  
 ; FILE REFERENCE: P-LJ 4852  
 ; CURRENT APPLICATION NUMBER: US/09/927,602  
 ; CURRENT FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: US 09/638,211  
 ; PRIOR FILING DATE: 2000-08-11  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 390  
 ; TYPE: PRF  
 ; ORGANISM: Homo Sapien



Mon Jan 13 09:14:03 2003

us-09-816-825-2.0115.rapb

Page 3

US-09-927-602-4

Query Match 4.9%; Score 19; DB 10; Length 390;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ACRSYSHVLEKVEFFNLQ 185  
|||||  
DB 165 ACRSYSHVLEKVEFFNLQ 183

RESULT 7

US-09-927-602-2  
Sequence 2, Application US/09927602  
Patent No. US20020061562A1  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
APPLICANT: Akama, Tomoya O.  
TITLE OF INVENTION: Methods of Treating Macular Corneal  
FILE REFERENCE: P-LJ 4852  
CURRENT APPLICATION NUMBER: US/09/927,602  
CURRENT FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: US 09/638,211  
PRIOR FILING DATE: 2000-08-11  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 395  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-927-602-2

Query Match 4.9%; Score 19; DB 10; Length 395;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSMRSGSFFVGOLF 62  
|||||  
DB 42 HVLVLSMRSGSFFVGOLF 60

RESULT 8

US-09-927-602-3  
Sequence 3, Application US/09927602  
Patent No. US20020061562A1  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
APPLICANT: Akama, Tomoya O.  
TITLE OF INVENTION: Methods of Treating Macular Corneal  
FILE REFERENCE: P-LJ 4852  
CURRENT APPLICATION NUMBER: US/09/927,602  
CURRENT FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: US 09/638,211  
PRIOR FILING DATE: 2000-08-11  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 395  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic construct  
NAME/KEY: VARIANT  
LOCATION: (1)...(395)  
OTHER INFORMATION: Xaa = any amino acid  
US-09-927-602-3

Query Match 4.9%; Score 19; DB 10; Length 395;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSMRSGSFFVGOLF 62  
|||||  
DB 42 HVLVLSMRSGSFFVGOLF 60

RESULT 9

US-09-927-602-5  
Sequence 5, Application US/09927602  
Patent No. US20020061562A1  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
APPLICANT: Akama, Tomoya O.  
TITLE OF INVENTION: Methods of Treating Macular Corneal  
FILE REFERENCE: P-LJ 4852  
CURRENT APPLICATION NUMBER: US/09/927,602  
CURRENT FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: US 09/638,211  
PRIOR FILING DATE: 2000-08-11  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 418  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-927-602-5

Query Match 4.9%; Score 19; DB 10; Length 418;  
Best Local Similarity 100.0%; Pred. No. 1.8e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSMRSGSFFVGOLF 62  
|||||  
DB 66 HVLVLSMRSGSFFVGOLF 84

Search completed: January 11, 2003, 01:32:30  
Job time : 14 secs

**This Page Blank (uspto)**

QY 130 RALCSAPACDIIIPODEIIPRACHRLCSQOPEVEYKACRSYSHVYLKEVFNLOSIV- 188  
 DB 1349 -----CTLLPQKDF-----ENLMLKIDKAMONDKILAIISVSELSLQ 1390  
 QY 189 -----PLKDPSPINLHIVLVRDPRVFRSRRRTKGDLM-----IDSRIVMQ 231  
 DB 1391 EIEMLKFRLESTDAQMGIRLKGISFQVFRQHFLLTKDLKALVENIDMTSLIL-- 1448  
 QY 232 HEQKLRKEDPPYVMQYICQSOLEIYKTIOSLPKALOERILLVRYEDLARAPVAGTSRMV 291  
 DB 1449 -----EYCSGFQDCCAVQLF----- 1465  
 QY 292 EFVGLFEPLPLQTVHNHITRGKMGDHAFTNAR 325  
 DB 1466 -----IETLLHNTNAGQCGDASMSAKR 1489

RESULT 15  
 ID TPSA\_CAEEL STANDARD; PRT; 380 AA.  
 AC 077081; Q9NEM9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-tyrosine sulfoltransferase A (EC 2.8.2.20) (Tyrosylprotein  
 sulfotransferase-A) (TPST-A).  
 GN Y11B2A.15.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=98406128; Pubmed=9733778;  
 RA Guyang Y.-B., Moore K.L.;  
 RT "Molecular cloning and expression of human and mouse tyrosylprotein  
 sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in  
 Caenorhabditis elegans.";  
 RL J. Biol. Chem. 273:24770-24774(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Sulston J.E.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP REVISIONS.  
 RA Durbin R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN  
 ACIDIC MOTIFS OF POLYPEPTIDES.  
 CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine =  
 adenosine 3',5'-bisphosphate + protein tyrosine-O-sulfate.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By  
 similarity).  
 CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF049709; AAC36062.1; -  
 DR EMBL; AL132904; CAC35844.1; -  
 DR WormPep; Y11B2A.15; CE26632.  
 DR InterPro; IPR000863; Sulfoltransferase.  
 DR Pfam; PF00685; Sulfoltransfer; 1.  
 KW Transferase; Transmembrane; Glycoprotein; Signal-anchor.  
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 28 380 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CAROHD 66 66 N-LINKED (GLCNAc...) (POTENTIAL).  
 SQ SEQUENCE 380 AA; 43313 MW; FF709BF0F1EDC95 CRC64;

Query Match 4.3%; Score 87.5; DB 1; Length 380;  
 Best Local Similarity 18.1%; Pred. No. 3.8;  
 Matches 79; Conservative 52; Mismatches 117; Indels 189; Gaps 19;

QY 5 KMKLLPLVVSOMALALFEHM-----YS-----HNSLS 35  
 DB 5 RELLVLPLV-----VFILFYPTARTADDPYSSNREKFNGAADGDESLPFHQISVR 60  
 QY 36 MKAQPERKHVLYLSSW--RSGSFFYGOLFQGHPDV-----FYLMEPAW-----HV 78  
 DB 61 SDDGYNRTSPFFIFGIVPRSGTTLRAMIDAHPEVRCCEETRVIPRIILNLSQMKKSEKE 120  
 QY 79 WMTFKQS--TAMMLHMAVDLIRAVFLCDMSYEDAYMEGPRGSSLPQWENSRAICAP 136  
 DB 121 WNRLOQAGVTGEVINNAISSFI-----MEIMVGHGDRAPR----- 155  
 QY 137 ADDIIPODEIIPRACHRLCSQOPEVEYKACRSYSHVYLKEVFNLOSIVPLKDPST 196  
 DB 156 -----LCNADPFTM-----KSAVYLKE-----LFP----- 175  
 QY 197 NLHIVLVRDPRVFRSRRRTKGDLMIDSRIV-----MGQHEQKLRKEDPPYVMQYIC 250  
 DB 176 NAKYLLMIRDSRATFNS-----IISRKVYIIGFDLNDRCQMTKWNNAIQIMVQDC 226  
 QY 251 QSOLEIYKTIOSLPKALOERILLVRYEDLARAPVAGTSRMVFEVLFLPLQTVHNHIT 310  
 DB 227 ESQVE-----KNCLEKYYEOLVLPHEQOMRRTIEEDIP----- 260  
 QY 311 RKGMGDHAFTNARDALNVSOAWMSLPEKVSRLQACGDMANLGYRHVRSDEQORN 370  
 DB 261 -----WD-----DKVLMHEQLIGDISLSNVR-SSQGVVVKP 291  
 QY 371 LLIDLISLV--TVPEQI 385  
 DB 292 VNIDALIKKWGTIPEDV 308

Search completed: January 11, 2003, 01:19:27  
 Job time : 39 secs

Best Local Similarity 20.7%; Pred. No. 24;  
Matches 71; Conservative 38; Mismatches 108; Indels 126; Gaps 18;

QY 23 FFMVSHNSSLKAKOPERMHVLVLSWRSQSSFGQLFGQHPDV-----68  
DB 367 FARLVHN-SSMCATIMEPOL-----KEFMGNWKGKMSVLAARSSVRAVCVN 414  
QY 69 -----FLMEPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVDAWMEPG 116  
DB 415 KFIKGLKPYSTFLNLNLTW--KWMENSTAMF-----FDTAEVD 453  
QY 117 -PROSSLFQWENS--RALSAPACDIIPODETIPRAHCLLC--SQQPEVEVER---166  
DB 454 VPEKLDLIFMGEGAGLVAAHITSRYVGTVP---LADRENMALLCMDSQKLIHMRFRMR 510  
QY 167 -ACRSYGHVYLKVRFPNLCSLYPLKDPNLHIVLVADPRAVFRSRETKGDLMTDS 225  
DB 511 GAWGAMHCVCISRE--FLLYVEARLKSSCL-----IAARRR-----545  
QY 226 RIVMGQHEOR-----LKKEDQPYVMQVICOSEIYKTIQSLPKALOERYLLVRE 277  
DB 546 ---GQKHEKLEAVEYGLKSSDALFRAMTYLCAARLEPWFSESGL-----RFFLTRGR 595  
QY 278 DLARAPVQTSRMEFGLFELPHLQTW--VANIT--RGKGM 315  
DB 596 NNIXGLNTNYTEGKRAVTGVNL---WSNVHEVSTFRKRM 633

## RESULT 13

F3ST\_FLABI STANDARD: PRT: 312 AA.

AC P52835;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Flavonol 3-sulfotransferase (EC 2.8.2.-) (F3-ST).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;  
OC Asteridae; easterids II; Asterales; Asteraceae; Asteroideae;  
OC Helianthaceae; Flaveria.  
OX NCBI\_Taxid=4224;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95083751; PubMed=7991681;  
RA Kharvornich S., Varin L., Gullick P., Ibrahim R.;  
RT "Cloning and regulation of flavonol 3-sulfotransferase in cell-  
suspension cultures of Flaveria bidentis.";  
RL Plant Physiol. 106:485-491(1994).  
CC - FUNCTION: CATALYZES THE SULFATE CONJUGATION OF QUERCETIN.  
CC RHAMNETIN AND ISORHAMNETIN BUT NOT KAEMPEROL. O-SULFURATION OF  
CC POSITION 3 OF FLAVONOL. MAY PLAY A ROLE IN AUXIN TRANSPORT.  
CC - SUBCELLULAR LOCATION: Cytoplasmic.  
CC - TISSUE SPECIFICITY: HIGHEST IN SHOOT TIPS AND LOWEST IN MATURE  
CC LEAVES AND ROOTS.  
CC - SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DB EMBL: U10275; AAA61638.1; -  
DR HSSP: P50224; ICDM;  
DR InterPro: IPR000863; Sulfotransferase.  
DR Pfam: PF00685; Sulfotransferase; 1.  
DR ProDom: PD001218; Sulfotransferase; 1.  
KW Transferase.  
FT BINDING 138 153 PAPS-BINDING SITE (BY SIMILARITY).  
SQ SEQUENCE 312 AA; 36457 MW; 29D174E8CB663FC2 CRC64;

Query Match 4.3%; Score 88; DB 1; Length 312;

Best Local Similarity 21.7%; Pred. No. 2.7;  
Matches 54; Conservative 41; Mismatches 86; Indels 68; Gaps 13;

QY 60 QLEQGHF-DVEFLMEPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVDAWMEPGPR 118  
DB 43 QIFKAPNPNVPLASV-----KSGTTLKALAFATITREKF-----78  
QY 119 ROSSLFQWENSRALCSPACDIIPODETIPRAHCLLC--SQQPEVEVERKACRSYHVLKE 178  
DB 79 -----DSS-----TSPLLTTPHD-CIP-----LL--EKDLEKIOENRNSLYTPIS- 117  
QY 179 VREFNLQSLYPLKDPNLHIVLVADPRAVFRSRETKGDLMTDSRIVMQHEQKILK 238  
DB 118 -THFHKSLPESAR--TSCKIYVITRNKKDYIVSYHFLRIV-----KLSV 162  
QY 239 EDQPY-YVMQVICO--SOLEIYKTIQSLPKALOER--YLLVREDLARAPVQTSRMY 291  
DB 163 EAPFEEAVDERCQGISGCPYWEHLIGYWKASLEKEPFIETLEKEDMKKDPVSVKRLA 222  
QY 292 EFVGLLEFLP 300  
DB 223 DFTGHPFTP 231

## RESULT 14

Y166\_HUMAN STANDARD: PRT: 2209 AA.

AC P50748;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein KIAA0166.  
GN KIAA0166.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=96281124; PubMed=8724849;  
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;  
RT Prediction of the coding sequences of unidentified human genes. V.  
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 3:17-24(1996).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DB EMBL: D79988; BA111483.1; -  
DR Genew; HGNC:17255; KNTC1.  
KW Hypothetical protein.  
SQ SEQUENCE 2209 AA; 250746 MW; 51AC948733CBFDAC CRC64;

Query Match 4.3%; Score 88; DB 1; Length 2209;

Best Local Similarity 16.2%; Pred. No. 33;  
Matches 54; Conservative 48; Mismatches 98; Indels 134; Gaps 10;

QY 29 HNISSLKAKOPERMHVLVLSWRSQSSFGQLFG--QHPDVYL-----M 72  
DB 1253 NISIALNLQNGS-----SQWELALREYVGSFGTCLOHSVSNMNAITLSEKLGERTTL 1305  
QY 73 EPAMVHWMTFKOSTAMMLHMAVRDLIRAVFLC---DASVPAYMEPGRROSSLQFQWENS 129  
DB 1306 VKSRHVVELKEKAVIFIRENATTLILHRVFNCRVLDLDAIGY-----1348

[illegible]

```

Db      269 YLLGQDPWMCAPGAAMWLTAOHSTRGGSSRW-----RE-----TCOMSA---LRP-311
OY      117 PRROSLFQWENSRALCSAPACDIIPOEETIFRAICRLIC-----SQGFEEVWE-165
          ||| |
Db      312 PRFGSSH-IYGARFL-----LLEEDFIPIFVALNMAMGGGSFSAGGGRKMSF---360
OY      226 RIVMGQHECKLKREDQPIFYVNOVICQSOLEIKTIOSLPKALOERYLLVRYDDLARAYVA-285
          ||| |
Db      411 FILMRITVDIYELEFAKTQLMMLMN-----LESRP-----VIEFDVRGYLA-454
OY      286 QTSRMVEEFGYLEFLRH-LQTWHNTIRGKMGDHAFHNARDALNVSQAWMSLYEKYS-344
          ||| |
Db      455 THSKR-----LPEHLCTILRN-----KPEDIKRVASKMLRGKRF-----488
OY      345 RLQKACGDANILLGRHVRSFEQDN-370
          ||| |
Db      489 -AVALGDIDTDLPTEHIQALLSRD-513

RESULT 12
RRPO_PWMR
ID      RRPO_PWMR STANDARD: PRT: 1968 AA.
AC      P17965; O89548;
DT      01-NOV-1990 (Rel. 16, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      RNA replication protein (147 kDa protei)n (ORF 1) [contains: RNA-
OS      directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OX      Potato virus M (strain Russian) (PVN).
XX      Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN      NCBI_Taxid=12168;
[1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-91116326; PubMed-1990070;
RA      Zavilev S.K., Kanyuka K.V., Levay K.E.;
RT      "The genome organization of potato virus M RNA.";
RL      J. Gen. Virol. 72:9-14(1991).
CC      - FUNCTION: RNA-replication. The central part of this protein
CC      possibly functions as an ATP-binding helicase.
CC      - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC      {RNA}(n).
CC      - SIMILARITY: CONTAINS 1 OTU DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@sib-sdb.ch).
CC      -----
DR      EMBL: D14449; BA003339.1; -.
DR      FIR: S21601; S21601.
DR      FIR: EN0093; EN0093.
DR      MEROPS: C23.001; -.
DR      InterPro: IPR003323; OTU.
DR      InterPro: IPR001788; RNA_dep_RNApol12.
DR      InterPro: IPR000606; Viral_helicase1.
DR      Pfam: PF009378; RNA_dep_RNApol12; 1.
DR      Pfam: PF01443; Viral_helicase1; 1.
DR      Pfam: PF02338; OTU; 1.
DR      PROSITE: PS50802; OTU; 1.
KW      ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW      transferase.
FT      DOMAIN 883 991 OTU.
FT      NP_BIND 1166 1173 ATP (POTENTIAL).
SQ      SEQUENCE 1968 AA; 223384 MW; 6F15A7991AD96AAC CRC64;
Query Match 4.4%; Score 89; DB 1; Length 1968;
```

Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A., Viari A., Wambuit R., Wedler E., Wedler H., Weitzengger T., Winters P., Wipst A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Zurcher A., "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*," Nature 390:249-256(1997).

SEQUENCE OF 1-1763 FROM N.A.

STRAIN-168 / PB1424:

Tognoni A., Grand G.

Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.

FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE SYNTHESIS OF A POLYPEPTIDE MOLECULE WHICH MAY BE INVOLVED IN SECONDARY METABOLISM.

COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES (POTENTIAL).

SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: Z99113; CAB13603.1; -

EMBL: Z35133; CAAB4505.1; -

Subtilisin: BG10931; pKSM.

Interpro: IPR002198; ADH\_short.

Interpro: IPR007794; Ketoacyl-synt.

Interpro: IPR001601; Methyltransf.

Interpro: IPR003880; Ppantn-attach.

Interpro: IPR000051; SAM\_bind.

Pfam: PF00106; adh\_short. 1.

Pfam: PF00109; ketoacyl-synt. 3.

Pfam: PF00550; pp-binding. 4.

Pfam: PF02801; ketoacyl-synt\_C. 3.

PROSITE: PS00012; PHOSPHOPANTETHEINE. 2.

PROSITE: PS00606; B\_KETOACYL\_SYNTHASE. 2.

PROSITE: PS50075; ACP\_DOMAIN. 4.

Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP; Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.

DOMAIN 295 364

ACTL CARRIER (ACP) 1.

BETA-KETOACYL SYNTHASE 1.

ACTL CARRIER (ACP) 2.

BETA-KETOACYL SYNTHASE 2.

BETA-KETOACYL SYNTHASE 3.

ACTL CARRIER (ACP) 3.

ACTL CARRIER (ACP) 4.

PHOSPHOPANTETHEINE (POTENTIAL).

PHOSPHOPANTETHEINE (POTENTIAL).

BETA-KETOACYL SYNTHASE (BY SIMILARITY).

PHOSPHOPANTETHEINE (POTENTIAL).

BETA-KETOACYL SYNTHASE (BY SIMILARITY).

PHOSPHOPANTETHEINE (POTENTIAL).

E -> V (IN REF. 2).

E -> E (IN REF. 2).

T -> S (IN REF. 2).

CONFICT 289

SEQUENCE 4273 AA: 477459 MW: 3886CFLA250AEB5A CRC64;

Query Match 4.4%; Score 89.5; DB 1; Length 4273;

Best Local Similarity 22.9%; Pred. No. 58;

Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

161 FEVVERACRSYSHVLTKEVFNQSLYPLKDP-----LNHIHLVDPRAVERSRER 216

36 YQVFOEHGVYQDELK-----NLTFYPLADESYDIALTIHVSREGETWSIILDCGK 90

217 TKGMLIDRSIVMGHE-QKIKKEDQPYVMQVYICSOLEIKTOSLPKALOERLLVR 275

91 QHESLSDKR-----QYETADMRKKEQTAFRESI-----DLNOMKSTADRLINLDEIYEOCR 142

276 YEDLAPAPAOQ-SRMYE-----FYGLEFPLHLOTWNIT--RCKMGDHAFTN 323

143 SOELVTHGMKKEGQYVEKBEAGVIDLAVGQALRHSAFLPHPLIDSGTGS----- 196

324 ARDALNVQAMRWNSLPEKVS---RLQKAC 350

197 --SCLISDQTWLPLYESFSASERLQKCC 224

RESULT 10

F4ST\_FLACH STANDARD: PRT; 320 AA.

AD P52837;

DT 01-OCT-1996 (rel. 34, Created)

DT 01-OCT-1996 (rel. 34, Last sequence update)

DT 01-OCT-1996 (rel. 34, Last annotation update)

DE Flavonol 4'-sulfoyltransferase (EC 2.8.2.-) (F4-ST).

OS *Flavaria chloraeifolia*.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Asteridae; eustersids II; Asterales; Asteraceae; Asteroideae; Helianthaceae; Flaveria.

OX NCBI\_TaxID=4228;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92159034; PubMed=1741382;

RA Varin L., Deluca V., Ibrahim R.K., Brissson N.;

RT "Molecular characterization of two plant flavonol sulfoyltransferases," Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).

RL [2]

RN PAPS-BINDING SITE.

RP MEDLINE=95279378; PubMed=7759495;

RX Varin L., Marsolais F., Brissson N.;

RT "Chimeric flavonol sulfoyltransferases define a domain responsible for substrate and position specificities," J. Biol. Chem. 270:12498-12502(1995).

RL [3]

RN FUNCTION: TRANSFERS SULFATE GROUP INTO FLAVONOL POSITION 4'.

CC MAY PLAY A ROLE IN AUXIN TRANSPORT.

CC SUBCELLULAR LOCATION: Cytoplasmic.

CC TISSUE SPECIFICITY: HIGHEST IN SHOOT TIPS AND LOWEST IN MATURE LEAVES AND ROOTS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE PLANT SULFOYLTRANSFERASE FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M84136; AAN33343.1; -

HSSP: P50224; ICDM.

Interpro: IPR000863; Sulfoyltransferase.

Pfam: PF00685; Sulfoyltransferase. 1.

ProDom: PD001218; Sulfoyltransferase; 1.

Transferrase.

BINDING 148 163

SEQUENCE 320 AA: 37255 MW: 68263659P6CCBCO CRC64;

Query Match 4.4%; Score 89; DB 1; Length 320;

Best Local Similarity 20.8%; Pred. No. 2.3;

Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12;

54 GSSFVGQLEFGQHP-DVETLMEPAWHVMTFQSTAMWIMHVRVIRAVFLCDMSVEDAY 112

47 GAILAQSGFKARPDVDFCSYP-----KSGTTLKALAVATVTR-----EKDFEF 91

113 MEP-----GPRQSLFO-WENSALCASAPACDIIPQDEIIPRACHCLSCQPF 162

92 TSPLLTNPANCIPYIEKDLKRIYENONNSCTP-----MATMPPYH 133

DR PROSITE: PS00410; DYNAMIN; 1.  
 KW Interferon induction; GTP-binding; Multigene family.  
 FT NP\_BIND 75 82 GTP (POTENTIAL).  
 FT NP\_BIND 176 180 GTP (POTENTIAL).  
 FT NP\_BIND 245 248 GTP (POTENTIAL).  
 SQ SEQUENCE 659 AA; 74951 MW; F55D63283BC865B3 CRC64;  
 Query Match 4.5%; Score 91; DB 1; Length 659;  
 Best Local Similarity 20.6%; Pred. No. 3.9;  
 Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;  
 QY 62 FGQHPDVFYLMERAMWMMFKOSTAMLMMAVRDLI---RAVELCDMSVDPAYVE--- 114  
 DB 301 FKHPDRALELDG-----KATVPCLARLTMELISHCKSLPLENOIKESHSTSE 353  
 QY 115 -----PGRROSSLFQWENSRA LCSAPACDIIPODEIPRACRLCS-QCPF--- 161  
 DB 354 ELQKGDADIPEDENKTLFLIEKINAF-NODITAVEGEIVREKCRFLFKRKEFFLM 412  
 QY 162 -EYERKACRSYSHVLKEVAFNLO-----SLYLKRPSTLNLH 200  
 DB 413 SEELERFQKSDALKYEVTFEMQYRGRELPGFVNYKTFENIRQIKLEPPAMEM--- 470  
 QY 201 VHLVRD-FRAVFR-----SRETKGDLMTDSRIYMGHEOKLKEDEPPYVMQV 248  
 DB 471 LHVTEIVRAFLTVSEKNSEFFNLRTKSKLEDIRL---EQITAEAKALHQMED 527  
 QY 249 ICOSLEIYKTIOSLPKALOERYLVRYEDL-----ARAPVAQTSNMYEVL 296  
 DB 528 IIVCODIVR-----KALOK---VREEAEEERKHGKRSQSPNQITSSMD----- 572  
 QY 297 EFLPHLOTW---VEN 308  
 DB 573 EIFOHLNAYROEARN 587  
 RESULT 8  
 C7D8-SOYBN STANDARD; PRT; 504 AA.  
 ID C7D8-SOYBN STANDARD; PRT; 504 AA.  
 AC 081974;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE Cytochrome P450 71D8 (EC 1.14.-.-) (P450 CPT7).  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OX NCBI\_TaxID=3647;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cy. Harosoy 63;  
 RX MEDLINE=98311068; PubMed=9648734;  
 RA Schopfer C.R., Ebel J.;  
 RT "Identification of elicitor-induced cytochrome P450s of soybean  
 (Glycine max L.) using differential display of mRNA."  
 RL Mol. Gen. Genet. 258:315-322(1998).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Y10493; CAAT7517.1; -  
 DR HSSP: P14779; 1UPZ.  
 DR Interf: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450; 1.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.

KW Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 444 444 HEME (BY SIMILARITY).  
 SQ SEQUENCE 504 AA; 57579 MW; 86E4FAEEC2FF2A6F CRC64;  
 Query Match 4.4%; Score 89.5; DB 1; Length 504;  
 Best Local Similarity 19.9%; Pred. No. 3.7;  
 Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;  
 QY 68 VFYLMERAMWMMFKOSTAMLMMAVRDLI RAVELCDMSVDPAYVEGPRROSSLPQWE 127  
 DB 14 VELLH--WLV-KTYKQKSSKL-----PGPMR----- 39  
 QY 128 NSRALCSAPACIITQDEIPRACRLCSQDPFEVAKCRSYS---HYVLKEVFNPL 184  
 DB 40 -----LP---IIGNLHQLAASLPDGLKLVKRGPLMHLDGLSTLIV 83  
 QY 185 QSLVPLKDPSTNLHIVLVRDRAVFRSREPTKGDLMDSRIYMGHEOKLKEDEPPY 244  
 DB 84 SS-FKMAEMKMTDHYHVQRQ-----LLAPQMYGARDIAFAPYGDYWR 129  
 QY 245 VMQVTCOSOLEIYKTIOSLPKALDE--RYLVRYSDLARAPVQTSNMYEVL----- 296  
 DB 130 QIRKICITLLELSAKRVQSFSHIRQDENKLLQSISSAGSPIDLSGLFSLGITVSPA 189  
 QY 297 -----EFL-----PHLOTWYHNIRKGMGDHFTNA----- 324  
 DB 190 FKENDDDDEPNSLVKRALITMTGFEVDMPFSLP-LHLTRQAKVHV-HQRAKIL 247  
 QY 325 RDALNVSQAWMSLPEKYSRLQKACGDAMNLGYRHSRQEQDQNLDDL 376  
 DB 248 EDLKRKM-----EKRTVKEGNG-----SEAEQDLVDVLL 279  
 RESULT 9  
 PKSM\_BACSU STANDARD; PRT; 4273 AA.  
 ID PKSM\_BACSU STANDARD; PRT; 4273 AA.  
 AC P40872; 031781;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 15-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2003 (Rel. 41, Last annotation update)  
 DE Putative polyketide synthase pksM.  
 DE PKSM OR PKSY.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168.  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertier M.G., Bessieres P., Bohloul A., Borcherdt S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,  
 RA Eutia K.D., Erington J., Fader C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Goldightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krohn S., Kumano M.,  
 RA Kurita K., Lapidus A., Laroche S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nock M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudea B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetle D., Porciliak S., Prescott A.M.,  
 RA Preece E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rochna E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
 RA Sekiguchi Y., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

Best Local Similarity 22.2%; Pred. No. 53;  
Matches 88; Conservative 66; Mismatches 128; Indels 115; Gaps 24;

QY 2 LPPKMKLLPLVYQMAILALFPMYSHNITSSLSMKAPREMHVLYVSWKSGSSPVGOL 61  
DB 1752 LLSQQNFILATQSAQA-----FLDQGHGHNLT-----PEEQQL-----QKIGEL 1792  
QY 62 FGQHPDVFLYLEPAWHVMTFKOSTAWMLHM-AVRDLIRAVFLCDSVFPAYMEPPFRQ 120  
DB 1793 KEQYS-----TSLAQSEALKOVOTLDELQK-FLDQHKFEFWSLE---RSE 1835  
QY 121 SSLQWENSRALGAPACDIIPODEIIPRAHCRLLCSQDF--EYVKAARSYSH----- 173  
DB 1836 KEL---ENMKKGGSSP-----ETLP-----SLKRGGSFSEDI-----SHKGLR 1873  
QY 174 -VLEKEVFFNLGSLYPLLDPSLNLIYHLVPRPRAVFRSREKTDLMIDSR----- 226  
DB 1874 FVTISGQKVLDMENSFKEGKEPS---EIGNLVKD-----KTKDATEHYTALHSKCPRLGS 1925  
QY 227 ---IVMGQHECKLKKEDOPYYVQVIGOSOLE--IKTISLPRALDERLLVRYEDLAR 281  
DB 1926 HLNLMLCOYHOFONSASLSQAMQA--CEANVEKLLSTVXASDPGLQEQ----- 1973  
QY 282 APVAQTSNMEYFVGLFELP--HLQTVVHNI--TRGKMGDHAFTNARDA-LNVSQAMRW 336  
DB 1974 --LATTKQLODELAHQVPEKLOKVARDIMEITGEPAQPHRHVQERTDLSLHPSQSLSY 2031  
QY 337 SLPEKYSRLQKACGDAMNLGTRHVSSEQORULL 373  
DB 2032 SLA-FRSSLQKALTAQS-----QSVQSELSLESL 2058

RESULT 6  
239F\_HUMAN STANDARD; PRT: 294 AA.  
AC Q15777;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Fetal brain protein 239 (239FB).  
GN C11ORF8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95080775; PubMed=7527372;  
RA Schwartz P., Neve R., Eisenman R., Gessler M., Bruns G.;  
RT "A WAGR region gene between PAX-6 and FSHB expressed in fetal brain.";  
RL Hum. Genet. 94:658-664(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96115606; PubMed=8666403;  
RA "Schwartz P., Eisenman R., Knoll J., Gessler M., Bruns G.;  
RT "cDNA sequence, genomic organization, and evolutionary conservation  
of a novel gene from the WAGR region.";  
RL Genomics 29:526-532(1995).  
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN FETAL BRAIN.  
CC -1- SIMILARITY: BELONGS TO THE UPF0046 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: U57911; AAC50564.1;  
CC Genew: HGNC:1180; C11orf8.  
CC MIM: 600911;  
CC InterPro: IPR004843; M-peptidase.

DR InterPro: IPR004844; S/T-phosphatase.  
DR Pfam: PF00149; Metallophos; 1.  
SQ SEQUENCE 294 AA: 33360 MW: 4392BC0DA1BFD1F0 CRC64;  
Query Match 4.5%; Score 91; DB 1; Length 294;  
Best Local Similarity 20.2%; Pred. No. 14;  
Matches 53; Conservative 30; Mismatches 89; Indels 90; Gaps 9;

QY 139 DLPDEIIPRAHCRLLCSQDFEVEVEKACRSYSHVL-----KEYRFFNLQSL 187  
DB 45 DPLPYDTPAGHTFRVCISDTHSRFDGIQMPYGDILLTGDFELGLPSEVKFRN----- 100  
QY 168 YPLKRPSSLNLIYHLVPRPRAVFRSREKTDLMIDSRIVM-GQHECKLKE-----D 240  
DB 101 -----DWLGNLPEYKVIAGNHELTFDEKFNADLVK 132  
QY 241 QPYVYVQVIGOSOLEIKYKISLPKALQERYLLVRYEDLARAPVQTSRM---YEVGL 296  
DB 133 QDYRFRFSYSLKLPDEFDVQSL-----LTNSIYLQDSYVYKGFRIYGA 177  
QY 297 EPLPLQTVHNIITRGKMGDHAFTNARDALNVSQAMRWSLPEKYSRLQKACGDAMNL 356  
DB 178 PMPWENGWGFNLPFGQSLLD-----KWNLDIPGIDILM-THGP--- 216  
QY 357 LGYRHYRSEQOORNLIDLST 378  
DB 217 LGFRMWPKELQRYGCVCELLNT 238

RESULT 7  
MX3\_RAT  
ID MX3\_RAT STANDARD; PRT: 659 AA.  
AC P18590;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Interferon-induced GMP-binding protein Mx3.  
GN MX3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91056598; PubMed=2173790;  
RA Meier E., Kunz G., Haller O., Arheiter H.;  
RT "Activity of rat Mx proteins against a rabdovirus.";  
RL J. Virol. 64:6263-6269(1990).  
CC -1- FUNCTION: DOES NOT SHOW ACTIVITY AGAINST INFLUENZA VIRUS OR VSV;  
CC ALTHOUGH IT ONLY DIFFERS FROM MX2 BY 8 POSITIONS.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- INDUCTION: BY INTERFERONS.  
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: X52713; CA36937.1;  
CC PIR: S11737; S11737.  
CC InterPro: IPR001401; Dynaminn.  
CC InterPro: IPR000375; Dynaminn\_central.  
CC InterPro: IPR003130; GED.  
CC Pfam: PF00350; dynaminn\_1.  
CC Pfam: PF01031; dynaminn\_2; 1.  
CC Pfam: PF02212; GED; 1.  
CC PRINTS: PR00195; DYNAMINN.  
CC SMART: SM00053; DYNc; 1.  
CC SMART: SM00302; GED; 1.



RT "Molecular cloning of macrophin, a human homologue of Drosophila  
 RT kakapo with a close structural similarity to plectin and dystrophin."  
 RL Biochem. Biophys. Res. Commun. 264:568-574(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20026884; PubMed=10559237;  
 RA Sun Y., Zhang J., Knaef S.K., Auclair D., Chang M.-S., Liu Y.,  
 RA Sutcliffe R., Salgia R., Griffin J.D., Ferland L.H., Chen L.B.;  
 RT "Molecular cloning and characterization of human trabeculin-alpha, a  
 RT giant protein defining a new family of actin-binding proteins."  
 RL J. Biol. Chem. 274:33522-33530(1999).  
 RN [3]  
 RP SEQUENCE OF 868-2350 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20039619; PubMed=10574462;  
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,  
 RA Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 6:337-345(1999).  
 RN [4]  
 RP SEQUENCE OF 1544-5057 FROM N.A.  
 RA Corby N.;  
 RT Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 3734-5430 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98116662; PubMed=9455484;  
 RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,  
 RA Nakajima D., Nomura N., Ohara O.;  
 RT "Characterization of cDNA clones in size-fractionated cDNA libraries  
 RT from human brain."  
 RL DNA Res. 4:345-349(1997).  
 CC -1- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-  
 CC LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO  
 CC MICROTUBULES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED.  
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB029290; BAA83821.1; -;  
 DR EMBL: AF141968; AAF06360.1; -;  
 DR EMBL: AB033077; BAA86565.1; -;  
 DR EMBL: AL137853; CAC1520.1; -;  
 DR EMBL: AB007934; BAA32310.1; -;  
 DR HSSP: Q01082; 1BKR.  
 DR Genew: HGNC:13664; MACF1.  
 DR Interpro: IPR001589; Actbind\_actu1n.  
 DR Interpro: IPR001715; Calponin-like.  
 DR Interpro: IPR002048; EF-hand.  
 DR Interpro: IPR003108; GAS2.  
 DR Interpro: IPR001452; SH3.  
 DR Interpro: IPR002017; Spectrin.  
 DR Pfam: PF00036; efhand; 2.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00435; Spectrin; 36.  
 DR Pfam: PF02187; GAS2; 1.  
 DR ProDom: PD000012; EF-hand; 1.  
 DR SMART: SM00033; CH; 2.  
 DR SMART: SM00054; Efn; 2.

DR SMART: SM00243; GAS2; 1.  
 DR SMART: SM00150; Spec; 35.  
 DR PROSITE: PS00019; ACTININ\_1; 1.  
 DR PROSITE: PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE: PS00021; CH; 2.  
 DR PROSITE: PS00018; EF\_HAND; 2.  
 DR PROSITE: PS00002; SH3; FALSE\_NEG.  
 KM Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.  
 FT DOMAIN 1  
 FT 78 181  
 FT 194 295  
 FT REPEAT 314 355  
 FT REPEAT 591 623  
 FT REPEAT 680 784  
 FT REPEAT 786 800  
 FT DOMAIN  
 FT 871 923  
 FT REPEAT 1250 1272  
 FT REPEAT 1287 1342  
 FT REPEAT 1455 1534  
 FT REPEAT 1547 1659  
 FT REPEAT 1815 1891  
 FT REPEAT 1932 2042  
 FT REPEAT 2260 2280  
 FT REPEAT 2372 2395  
 FT REPEAT 2398 2507  
 FT REPEAT 2510 2618  
 FT REPEAT 2621 2728  
 FT REPEAT 2731 2838  
 FT REPEAT 2841 2945  
 FT REPEAT 2967 3024  
 FT REPEAT 3136 3163  
 FT REPEAT 3187 3274  
 FT REPEAT 3277 3383  
 FT REPEAT 3386 3492  
 FT REPEAT 3495 3601  
 FT REPEAT 3604 3673  
 FT REPEAT 3713 3819  
 FT REPEAT 3832 3927  
 FT REPEAT 3982 4043  
 FT REPEAT 4046 4152  
 FT REPEAT 4155 4262  
 FT REPEAT 4265 4371  
 FT REPEAT 4374 4481  
 FT REPEAT 4484 4590  
 FT REPEAT 4593 4700  
 FT REPEAT 4707 4808  
 FT REPEAT 4811 4917  
 FT REPEAT 4920 4985  
 FT REPEAT 5030 5054  
 FT CA\_BIND 5096 5107  
 FT CA\_BIND 5132 5143  
 FT DOMAIN 5276 5283  
 FT DOMAIN 5355 5370  
 FT CONFLICT 1 72  
 FT 575 594  
 FT CONFLICT 1487 1487  
 FT CONFLICT 1963 1963  
 FT CONFLICT 2052 2052  
 FT CONFLICT 2083 2083  
 FT CONFLICT 2250 2250  
 FT CONFLICT 2321 2321  
 FT CONFLICT 2343 2363  
 FT CONFLICT 2344 2350  
 FT CONFLICT 2523 2523  
 FT CONFLICT 4670 4670  
 FT CONFLICT 4833 4833  
 SQ SEQUENCE 5430 AA; 620346 MW; 91ADB7E7590B440B CRC64;  
 Query Match 4.5%; Score 91.5; DB 1; Length 5430;  
 VAISSEDEGNIREYVELLS -> GPSAPLKKKAISDLCMN  
 YCL (IN REF. 1).  
 A -> T (IN REF. 2 AND 3).  
 V -> A (IN REF. 1).  
 E -> D (IN REF. 1).  
 E -> K (IN REF. 2).  
 M -> Y (IN REF. 2).  
 C -> Y (IN REF. 1).  
 MISSING (IN REF. 2).  
 SLIPSVG -> EYRLFKI (IN REF. 3).  
 O -> R (IN REF. 2 AND 4).  
 S -> T (IN REF. 2).  
 MISSING (IN REF. 2).  
 4 X 4 AA TANDEN REPEATS OF [GS]-S-R [AR].  
 MSSDEFTLSRSCSRSRSRSGYSRSGSLSPGPD  
 TLPMPLPLEHKKRSQSDVDPAPRAVAV -> MPPLVLM  
 AGIPRGVSGSLPILPGKPKCTASRAVAVI (IN REF.  
 2).

```

CC -1- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.
CC PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO THE FUCOSE
CC OF THE NOD FACTOR.
CC -1- SIMILARITY: LIMITED TO NODH AND TO C. ELEGANS F42G9.8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y09415; CAA70569.1; -
DR EMBL: AE000076; AAB91690.1; -
KW Nucleation; Transferrase; Plasmid.
SQ SEQUENCE 419 AA; 46569 MW; 848C48E0416AA1F CRC64;

Query Match
Best Local Similarity 21.5%; Score 104.5; DB 1; Length 419;
Matches 78; Conservative 48; Mismatches 134; Indels 103; Gaps 18;

QY 34 LSKRAQPEPMHVLVLSMRSGSSFVGLFSGQHPDVEFLMEPMHWMTFKOSTAMMLHMA 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MSRPVSPDLPLICPLIGIPRSQTTLLAHILQOHDPITAPPEP-----WL----- 43
QY 94 VRDILRAVFLCDMSVPAYMEPRQSSSLFQWENSRLCSAPACDIIPODEII----- 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 44 -----MLALAFGRVDRHRA-----GASLLQVACDELGIIDRLSVSRVF 84
QY 148 -PRAHCRLCSQPFVEYKACRSYSHVLEKEVFENLQSLPLKDPNSLNIHVLVRD 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 85 ADAAYSGVLAAGKRLTIDKTPRYM--MVL-----YLHSLY--EAP-----HIL-LLRN 130
QY 207 PRAFRSREPRKGLMIDSRIVMGQHKLKEQRYVWQVYQSQ-----LE 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 131 PYATAALSKSTGWGVPFSEKCP-----PTVSCLALVYCTPTAAVALA 174
QY 256 IYKTIQSLPALQER-----YLVAIEDLARAFAVQTSRM-----YFEGLEFLPHQTWV 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 175 LADVLGLPFLAMQGRHRHGYVAYERLVERPDEIQRVLAGLGYDPAGIVFAGVEGT-- 232
QY 307 HNTIRGKGMD-HAFHTINARDALNVSOAKWSLPEKVSRLQKACG-DAMNLLGY-----R 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 233 -EYLRSSFGRRLKRAVDNRSV-ETWRTETLIEEMQVTVDLVGADLLVELGYEQSLQ 290
QY 361 HVR 363
   : :
DB 291 HAR 293

RESULT 4
TPSB_CAEEL
ID TPSB_CAEEL STANDARD: PRT; 359 AA.
AC 020351;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative protein-tyrosine sulfotransferase (EC 2.8.2.20)
DE (Tyrosyl)protein sulfotransferase (TPST).
GN F42G9.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Talc A.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
CC ACIDIC MOTIFS OF POLYPEPTIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine =

```

```

CC adenosine 3',5'-diphosphate + protein tyrosine-O-sulfate.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00051; AAB91354.1; -
DR WormPeP: F42G9.8; CE07235.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer; 1.
KW Hypothetical protein; Transferrase; Transmembrane; Glycoprotein;
KW Signal-anchor.
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
SQ SEQUENCE 359 AA; 41469 MW; 9FC2F44539B8CC53 CRC64;

Query Match
Best Local Similarity 19.5%; Score 92; DB 1; Length 359;
Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;

QY 52 RGGSSFVGLFSGQHPDV---FLMEPMHWMTFKOSTAMMLHNAVBDLRAVFLCDMS 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 95 RSGTILMARILDAHPDVRGCGETMLPSFLWQAQWRND-WYNSNGI-----TQE 143
QY 108 VPDVAYMEPRRQSSSLFQWENSRLCSAPACDIIPQ-DEIIPRAHCRLCSQPFVEYK 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 144 VPDVAY-----SAFTETIVAKHSLAR-----LCKNDP----- 172
QY 167 ACRSYSHVLEKEVFENLQSLPLKDPNSLNIHVLVRDPRVRSERK---GGLM 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 ---YATMLPLTR-----RLXP-----NAKFTIMEDARAVAHSMIERKVPAGYNT 216
QY 223 ISRLIVMGQHKLKEKDEPPYVWQVYQSQ-----LEIKTIQSLPKLQEKYLLVRD 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 217 SPEISMEVWQNDLKR-----MTFQCNNAPOGQIKY-----YER 251
QY 279 LARAFAVQTSRMVEFGLEFLPHLQTWVHNTIRGKGMDHAFH-TNARDALNVSOAKRWS 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 252 LIQKPAEELRLITNFDLPFSQMLRHQDLIGDEVLDNDQERSASQYKNSITFKALTSWF 311
QY 338 LPT-EKVSRLQKACGDAMNLLGY 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 312 DCFSEETLRKLDVAPFLGILGY 334

RESULT 5
ACF7_HUMAN
ID ACF7_HUMAN STANDARD: PRT; 5430 AA.
AC Q90PN3; Q90KPO; Q9ULG9; Q9H540; Q75053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (Microphin) (Tribeculin-alpha)
DE (620 kDa actin-binding protein) (ABP620).
GN ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20001959; Pubmed=10529403;
RA Okada T., Matsuda S., Nakatsugawa S., Ichigotani Y., Iwahashi N.,
Takahashi M., Ishigaki T., Hamaguchi M.;

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 21:45:42 : Search time 22 Seconds  
(without alignments)  
727.721 Million cell updates/sec

Title: US-09-816-825-2

Sequence: 1 MLPPKMKLLFLVFSQMAIL.....EQRLNLLDLSTWVEQIR 386

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500.5	24.6	458	1	C6ST_CHICK
2	116	5.7	388	1	YM67_MYCTU
3	104.5	5.1	419	1	NOE6_RHISN
4	92	4.5	359	1	TPSH_CAEEL
5	91.5	4.5	5430	1	ACFT_HUMAN
6	91	4.5	294	1	239F_HUMAN
7	91	4.5	659	1	MX3_RAT
8	89.5	4.4	504	1	C7D8_SOYBN
9	89.5	4.4	320	1	PKSM_BACSU
10	89	4.4	524	1	MPRA_RAT
11	89	4.4	320	1	F4ST_FLACH
12	89	4.4	1968	1	RREO_PYMR
13	88	4.3	312	1	F3ST_FLABI
14	88	4.3	2209	1	Y166_HUMAN
15	87.5	4.3	380	1	TPSA_CAEEL
16	87	4.3	940	1	CHRD_BRARE
17	86.5	4.2	748	1	MURA_MOUSE
18	86.5	4.2	831	1	DPOF_THETH
19	86	4.2	823	1	SCH9_YRASI
20	85	4.2	632	1	YK3_CAEEL
21	85	4.2	1237	1	YD2_SCHPO
22	85	4.2	2688	1	ZEP1_MOUSE
23	84.5	4.1	2493	1	YB4_YEAST
24	83.5	4.1	864	1	GLND_PASMU
25	83	4.1	222	1	NOCG_XENLA
26	82.5	4.0	2549	1	FRAB_HUMAN
27	82.5	4.0	2549	1	FRAP_MOUSE
28	82.5	4.0	2549	1	FRAP_MOUSE
29	82	4.0	357	1	BUR_THETN
30	82	4.0	2352	1	MOKC_SCHPO
31	81.5	4.0	309	1	F3ST_FLABI
32	81.5	4.0	311	1	F3ST_FLACH
33	81.5	4.0	2109	1	RRLP_VSVCO

34	81	4.0	534	1	THAS_PIG	P47787 sus scrofa
35	81	4.0	659	1	MX2_RAT	P18589 rattus norv
36	80.5	3.9	223	1	NOCG_CHICK	O93525 gallus gall
37	80.5	3.9	417	1	IB53_HSVSA	P13193 herpesviru
38	80.5	3.9	503	1	PODX_MOUSE	O910M4 mus musculu
39	80.5	3.9	705	1	XVAL_BACST	P45702 bacillus st
40	80.5	3.9	884	1	SECA_PORPU	P51381 porphyra pu
41	80	3.9	379	1	VERM_DROME	P20351 crosophila
42	80	3.9	475	1	SYC_CHLMT	O99160 chlamydia m
43	80	3.9	604	1	SR68_DROME	O99582 crosophila
44	80	3.9	977	1	KEMS_MOUSE	P09581 mus musculu
45	80	3.9	978	1	KEMS_RAT	O00495 rattus norv

## ALIGNMENTS

RESULT 1	CSST_CHICK	STANDARD:	PRT:	458 AA.
AC	C6ST_CHICK			
ID	O92179;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OK	NCBI-TaxID=9031;			
RN	(1)			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.			
RC	STRAIN=White leghorn; TISSUE=Embryonic chondrocytes;			
RX	MEDLINE=95355490; PubMed=7629189;			
RA	Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,			
RA	Shinomura T., Habuchi O.;			
RT	"Molecular cloning and expression of chick chondrocyte chondroitin 6-			
RT	sulfotransferase."			
RL	J. Biol. Chem. 270:18575-18580(1995).			
CC	- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN			
CC	SULFATE.			
CC	- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin =			
CC	adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.			
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE			
CC	(BY SIMILARITY).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL: D49915; BAA0655.1; "			
KW	Transferrase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.			
FT	DOMAIN 1 23			
FT	DOMAIN 1 23			
FT	TRANSMEM 24 37			
FT	DOMAIN 38 458			
FT	CARBOHYD 62 62			
FT	CARBOHYD 73 73			
FT	CARBOHYD 95 95			
FT	CARBOHYD 236 236			
FT	CARBOHYD 399 399			
FT	CARBOHYD 443 443			
FT	SEQUENCE 458 AA: 52253 MW; G9A3B7D0A5086F0C CRC64;			
SQ	-----			
Query Match	24.6%; Score 500.5; DB 1; Length 458;			
Best Local Similarity	33.2%; Pred. No. 2.2e-35;			
Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;				
34	LSMKQPEPMHVLVLSNRSGSFGQLFGCHPDVFTYLMPEPMH--WMTKQSTAMMLH 91			
1				

## RESULT 15

C69679

polyketide synthase pksM - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 03-Nov-2000

C:Accession: C69679

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berber  
C.; Bron, S.; Broutelle, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A.; Mathers, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y., M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A.; Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akuch, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MIMD:98044053; PMID:9384377

A:Accession: C69679

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-4273 &lt;KUN&gt;

A:Cross-references: GB:Z59113; GB:AL009126; NID:G2634090; PIDN:CA13603.1; PID:G2634103

A:Experimental source: strain 168

C:Genetics:

A:Gene: pksM

C:Superfamily: Bacillus subtilis polyketide synthase pksM; 3-oxoacyl-[acyl]-carrier-prote

C:Keywords: carrier protein

F:293-363/Domain: acyl carrier protein homology &lt;ACPI&gt;

F:414-829/Domain: 3-oxoacyl-[acyl]-carrier-protein synthase I homology &lt;OASI&gt;

F:2340-2732/Domain: 3-oxoacyl-[acyl]-carrier-protein synthase I homology &lt;OAS&gt;

F:3145-3320/Domain: short-chain alcohol dehydrogenase homology &lt;SADI&gt;

F:3550-3942/Domain: 3-oxoacyl-[acyl]-carrier-protein synthase I homology &lt;OAS3&gt;

F:4138-4208/Domain: acyl carrier protein homology &lt;ACPI&gt;

Query Match

Best Local Similarity 4.4%; Score 89.5; DB 2; Length 4273;

Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

```
QY 161 FEVEKACRSYSHVLEKVRFFNLQSLYPLKDPSP---LNLHIVLVRDPRAVFRSRER 216
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 36 YQVQEHGAYVQDELEK-----NLTIFFPLIADSEYDIALTIHVSEREGSTWIIIDGOK 90
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 217 TKGDLMTDSRLVMQGH-QKLKEDDPYTYMQVYCOSQLEITYKTIOSLPKALDERILVR 275
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 91 QHGESLSDKR---QYETA DMHREQTAFASFI---DLNQMSIADRIINLDEIYEOCR 142
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 276 YEDLARAPVQOT-SRMEY-----FVGLLEFLPHQTQVHNNT--RKGMDHAFHTN 323
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 143 SQELVHTGMKAKESQITEAKGAVIDLAVQGFALRHSDAFLPHPTLIDSGSIGS----- 196
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 324 ARDALNVQAARMSLPEYKVS---RLQKAC 350
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 197 --SCLISDQTMVLPFLYESEFSASERLQKGC 224
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

Search completed: January 11, 2003, 01:21:55  
Job time : 54 secs

**This Page Blank (uspto)**



Qy 340 YKVSRLQKACGDAMNLGYRHVRSQ-----EQRLILLDLSTWTVPEOI 385  
 Db 1344 KLVSRKSYNGEYARLSTYILEGDEKSQLLEQFTFIVEGSLRDPDSV 1398

## RESULT 9

T16350

hypothetical protein F42G9.8 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C&gt;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000

R:Taich, A.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of *C. elegans* cosmid F42G9.

A:Reference number: Z18498

A:Accession: T16350

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-359 &lt;TAI&gt;

A:Cross-references: EMBL:000051; NID:g1216305; PID:g1216308; PID:AAA91354.1; CESP:F42G9

A:Experimental source: strain Bristol N2

A:Gene: CESP:F42G9.8

A:Mutons: 26/2; 50/2; 113/2; 155/3; 220/3; 290/2

C:Superfamily: *Caenorhabditis elegans* hypothetical protein F42G9.8

Query Match 4.5%; Score 92; DB 2; Length 359;  
 Best Local Similarity 19.5%; Pred. No. 3.8;  
 Matches 63; Conservative 48; Mismatches 114; Indels 96; Gaps 15;

Qy 52 RSGSSTFGQFQHPDV---FLMEPAHVMVTFKOSTAMLMHMAVRDILRAVFLCDMS 107  
 Db 95 RSGTTLRALIDAHDPVRCGCTMLPSPFLTQAGWRND-VVNNSGI-----TQE 143  
 Qy 108 VFDAVMEPGFRQSSLEFQWENSRLCSAPACDIIPQ-DEIPRAHCLLCSQPFVEVER 166  
 Db 144 VFDDAV-----SAFTTEIVAAHSELAPE-----LCKNP----- 172  
 Qy 167 ACSRSNVVLEKEVFENLQSLVPLKDPSLNLIHVLVDRPRAVRSRRTK-----GDLM 222  
 Db 173 ---YALMLPTIR-----RLYP-----NAKFIIMIRDAVHSMIEKRVAVGYNM 216  
 Qy 223 IDSRIVAGQHEOKLKKEDQYVYMOYICQSQ-----LEIKTIOSLKALOERILVRYED 278  
 Db 217 SDELSMVCNQMLERK-----MTQCNNAPQCICKVY-----YER 251  
 Qy 279 LARAFAQTSRMVEFYGLEFLPELQTVNHTTGKMGCDHAFH-TNARDALNVSQAMRS 337  
 Db 252 LIQKPAEEIIRITNFDLPFSQOMLRHODLIGDEVLDNQDESSASQVKSINWKALTSNF 311  
 Qy 338 LPY-EKVSRLQKACGDAMNLGY 359  
 Db 312 DCFSEETLRKLDVAPFLIGLY 334

## RESULT 10

JEO196

hydroxysteroid sulfotransferase (EC 2.8.-.-) - mouse

C:Species: *Mus musculus* (house mouse)

C&gt;Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 18-Jun-1999

C:Accession: JEO196

R:Sakakibara, Y.; Yanagisawa, K.; Takami, Y.; Nakayama, T.; Saito, M.; Liu, M.C.

Biochem. Biophys. Res. Commun. 247, 681-686, 1998

A&gt;Title: Molecular cloning, expression, and functional characterization of novel mouse s

A:Reference number: JEO196; M01D:98521187; PMID:9647753

A:Accession: JEO196

A:Molecule type: mRNA

A:Residues: 1-338 &lt;SAK&gt;

A:Cross-references: GR:AF026072; NID:q2570897; PIDN:AA669918.1; PID:q2570898

C:Comment: This enzyme catalyzes the transfer of a sulfonate group from the active sulfat

C:Superfamily: alcohol sulfotransferase

C:Keywords: transferase

Query Match 4.5%; Score 91.5; DB 2; Length 338;  
 Best Local Similarity 22.2%; Pred. No. 3.9;  
 Matches 52; Conservative 32; Mismatches 79; Indels 71; Gaps 11;

Qy 67 DVEYLMEPAHVMVTFKOSTAMLMHMAVRDILRAVFLCDMSVFDVMEPGFRQSSLEFQW 126  
 Db 59 DIFIVYIP-----KSGTMMHTE-----IVCLIKND-----GDSWTRSEPIWO- 96  
 Qy 127 ENSRALCSAPACD-IIPQDEIIPRAHCLLCSQPFVEVERKACRSYSHVLEKEVFENLQ 185  
 Db 97 -----RAPMCETIISAFNVLDPSPRIMSSHLPIELFTKA-----FF--- 133  
 Qy 186 SLVPLKDPSSLNLIHVLVDRPRAVRSRRTKQDLMIDSRIVMGQHQKIKKEDQPIYV 245  
 Db 134 -----SSRAKVIYVGNRPDVVVS-----LYYSKITA-GQ-----LKDPTMPQOF 172  
 Qy 246 MQVICOQLIEIKTIOSLKAL-----QERYLLVREYEDLARAFAQTSRMVEFYV 295  
 Db 173 LQNFLKGEVQGSFWDPIKQWIRMQNQEFLPITYEELQDQDLRSVQKICERTLG 226

## RESULT 11

S11737

resistance protein Mx3, interferon-regulated - rat

C:Species: *Rattus norvegicus* (Norway rat)

C&gt;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Mar-2000

C:Accession: S11737

R:Meier, E.; Kunz, G.; Haller, O.; Arnheiter, H.

submitted to the EMBL Data Library, April 1990

A:Description: Unexpected activity of rat Mx proteins against a Rhabdovirus.

A:Reference number: S11735

A:Accession: S11737

A:Molecule type: mRNA

A:Residues: 1-659 &lt;EMBL&gt;

A:Cross-references: EMBL:X52713; NID:q56724; PIDN:CAA36937.1; PID:q56725

C:Superfamily: dynamin-related protein vps1

Query Match 4.5%; Score 91; DB 2; Length 659;  
 Best Local Similarity 20.6%; Pred. No. 10;  
 Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;

Qy 62 PQGHDPVYLMEPAHVMVTFKOSTAMLMHMAVRDIL---RAVFLCDMSVFDVME-- 114  
 Db 301 FKHEPQFALLLEDG-----KATVPCLAEIRLTMLSHICKSLPLENQIKESHQSTSE 353  
 Qy 115 -----PGRROSSLFQWENSRLCSAPACDIIPQDEIIPRAHCLLCSQPFVEVER 161  
 Db 354 ELQKYGADIPEDENKTLFLEKINAF-NQDITAVEGEELVREKCRFLTKLKKRFFLV 412  
 Qy 162 -EVEKACRSYSHVLEKEVFENLQ-----STVPLKDPSSLNLI 200  
 Db 413 SEEIRNFQKSDALYKEVTFEMQYRGRELPGFVNYKTFENITIRQIKTLEPAMEH-- 470  
 Qy 201 VHLVDR-PRAYR-----SRERTKGDLMIDSRIVMGQHQKIKKEDQPIYVQV 248  
 Db 471 LHKVTEIYRAAFTVSEKNESEFNLHRTTKSKLEDLRL--EDQTEAKKIRLHFMQ 527  
 Qy 249 ICOSQLEIYKTIQSLKALOERYLLVRYEDL-----ARAFAQTSRMVEFYVGL 296  
 Db 528 LIYQDDQIYR-----KALQK-----VREEEADEEERKHGKRSQSNLQTSMD 572  
 Qy 297 EFLPHLQTV---VHN 308  
 Db 573 EIFQHLNAYROEAAH 587

## RESULT 12

E64477

replication factor C homolog - *Methanococcus jannaschii*C:Species: *Methanococcus jannaschii*

C&gt;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C:Accession: E64477



A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier, J.; Harwood, C.R.; Hentzel, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulo, M.F.; Koehler, P.; Kohnstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, Rieger, M.; Rivolta, C.; Rooha, E.; Roche, B.; Rose, M.; Sadle, Y.; Sato, T.; Scanlon, A.; Authors: Schelch, S.; Schoeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Authors: M. Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A65580; MUID:98044033; PMID:9384377

A:Accession: B69718

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3588 <RNN>

A:Cross-references: GB:299105; GB:AL009126; NID:92632457; PIDN:CAB12142.1; PID:g2632634

A:Experimental source: strain 168

R:Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M.M.; Zuber, P.; Yamane, K.

Nucleic Acids Res. 21, 93-97, 1993

A:Title: Nucleotide sequence of 5' portion of *srfa* that contains the region required for

A:Reference number: S35517; MUID:93181186; PMID:8441623

A:Accession: S35517

A:Status: nucleic acid sequence not shown; significant sequence differences

A:Molecule type: DNA

A:Cross-references: EMBL:D13262; NID:q216345; PID:g216345

A:Experimental source: strain 168 tpc2

A:Note: protein sequence not complete, the nucleotide sequence was submitted to the EMBL

J.Nakano, M.M.; Manuison, R.; Myers, A.; Curry, J.; Grossman, A.D.; Zuber, P.

J. Bacteriol. 173, 1770-1778, 1991

A:Title: *srfa* is an operon required for surfactin production, competence development, an

A:Reference number: A37323; MUID:91154134; PMID:1847909

A:Accession: A37323

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-46; 'EV', 49-145, 'T', 147-150, 'L', 152, 'AN', 155-280, 'T', 282-307, 'SF', 310-384

A:Cross-references: GB:M59339; NID:914366

J.Nakano, M.M.; Xia, L.; Zuber, P.

J. Bacteriol. 173, 5487-5493, 1991

A:Title: transcription initiation region of the *srfa* operon, which is controlled by the

A:Reference number: A43705; MUID:9158326; PMID:1715856

A:Contents: annotation

R:Fabriz, C.; Quentin, Y.; Guisepi, A.; Busuttill, J.; Haeck, J.; Denicot, F.

submitted to the EMBL Data Library, March 1993

A:Reference number: S46967

A:Accession: S46967

A:Molecule type: DNA

A:Residues: 3249-3271, 'A', 3273-3316, 'R', 3318-3451, 'V', 3453-3483, 'DE', 3486-3487, 'DAGL', 34

A:Cross-references: EMBL:X72672; NID:9516358; PIDN:CA51222.1; PID:9516359

A:Experimental source: strain 168 tpc2

C:Comment: This enzyme is one of several in the multienzyme complex that synthesizes the

C:Genetics:

A:Gene: *srfaA*

A:Note: *srfaA* is the first gene of the *srfa* operon

C:Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h

C:Keywords: antibiotic biosynthesis; carrier protein; phosphopantetheine; phosphoprotein

F:507-956/Domain: acetate-CoA ligase homology <ACLI>

F:974-1042/Domain: acyl carrier protein homology <ACP1>

F:1043-1488/Domain: repeat <RPT1>

F:1549-1993/Domain: acetate-CoA ligase homology <ACLI2>

F:2011-2079/Domain: acyl carrier protein homology <ACP2>

F:2080-2537/Domain: repeat <RPT2>

F:2589-3053/Domain: acetate-CoA ligase homology <ACP3>

F:3042-3109/Domain: acyl carrier protein homology <ACP3>

F:1006,2043,3074/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 4.7%; Score 95; DB 2; Length 3588;

Best Local Similarity 19.7%; Pred. No. 46;

Matches 63; Conservative 48; Mismatches 111; Indels 98; Gaps 13;

14 VSQMAIATLFFHMYSHNSSLKAKQPERNHVILVLSWRSQSSFGVQLFGQHPDVFLME 73

Db 990 VERAGTIDNEPFTGSHLKLTKLTKH-----ETGIEIPQGFIEHPTITLAE 1040

QY 74 PAHVMTFKOSTAMLMHAYRDLIRAVFLCDMSVEDAYVEPGRROSSLFQENSRALC 133

Db 1041 EADH-----RESKAF-----AVIEPAEKDEHYPLMHSSSEHS 1073

QY 134 SAPA-----CDIIEQDEII-----PRACRLTCS 157

Db 1074 SASSRMRESAIIHSSSSSEGRRYKAGARISGIVPTPVLETSTVLNENSTPROKIH-VCV 1132

QY 158 QQPEVEKACRSYSHVYLKEVFPFQISYPLDKDSL-----NLHIV-----HLVDP 208

Db 1133 DFNEMIRGCRSDALHSSVPTFDLAKA-PLFRIGLLGLEENRMLFDMHLLSDGV 1191

QY 209 AVFSRERTKQDLMD--SRIVMGQHECKLKEQDPYVWQVYQSQLEIYKTQSLPKA 266

Db 1192 SI-----GIMLEELARIYKGEQLPDLRLQKYDAVWQ--SRQAEGYKKDQAYWKE 1240

QY 267 L--QERYLVREEDLAPV 284

Db 1241 VFAGEPLVQLISDYPPEPV 1260

RESULT 8

113288

mei-41 protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 13-Aug-1999 #sequence, revision 13-Aug-1999 #text, change 17-Nov-2000

C:Accession: T13288

R:Harl, K.L.; Santerre, A.; Sekelsky, J.J.; McKim, K.S.; Boyd, J.B.; Hawley, R.S.

Cell 82, 815-821, 1995

A:Title: The mei-41 gene of *D. melanogaster* is a structural and functional homolog of

A:Reference number: 211072; MUID:95401271; PMID:7671309

A:Accession: T13288

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2354 <HAR>

A:Cross-references: EMBL:U34925; NID:9998351; PID:9998353; PIDN:AAC46881.1

C:Genetics:

A:Gene: mei-41

A:Cross-references: FlyBase:FBgn0004367

A:Introns: 650/3; 748/3; 2313/3

C:Function:

A:Description: involved in cell cycle checkpoint and meiotic recombination

Query Match 4.5%; Score 92.5; DB 2; Length 2354;

Best Local Similarity 20.5%; Pred. No. 43;

Matches 85; Conservative 52; Mismatches 147; Indels 131; Gaps 20;

28 SHNLS---LSMKAPERNHIVLSSWRSQSSFGVQLFGQHPDVFLMEPAHVMTFKO 84

Db 1058 NYFNASPOHLPLNLSDFALVATLSLGRYQF--QONTKHVDSFL----- 1102

QY 85 STAMMLHAYRDLIRAVFLCDMSVEDAYMEPGRROSSLFQENSRALCSAPACDIIPOD 144

Db 1103 -----AIGCTLA-----ICGTS-----PRQKKVQLMQ-----SLPA---RMRQ 1133

QY 145 EIIIPRACHLCSQQPEVEKAC--RSYSHVYLKEVRF-----NQS 186

Db 1134 LMEPMHSCYTCVHRPSCLOQPLFGSHSYHYEEMAFVMSRLIDVLPSSGRHLISS 1193

QY 187 LVPLKLD-----PSLNLHIV-----HEVRPRAVFSRERTKGLMDSRIY 228

Db 1194 YKPCIKRDSNMSTYRPIYLLHALLLECTEQGNHIOEERMAVLAQNES-----SSSV 1246

QY 229 MOQHECKLKEQDPYVWQVYQSQLEIYKTQSLPKALQERYLLVRVEDLARAVAOVS 288

Db 1247 RGOEELGAIKEN-----AFKOFESRKYAAGI-KPLASTLVDRKEDSSRV----- 1291

QY 289 RMYEYVGLFELHLOTW-----VHNITRGKMGDHAFTNAR--DALNVQAWRWSLP 339

Db 1292 RLAGKCAELDLDFLQRLNLEWRQRIHGRSTG-GKPEPTISNVRKIHEPLNL-----IP 1343

Db 389 DSNQITSTQKNSRQFEKWMFSTPFLAQVYQDACEPAKRLFGTKLASSAQEDLTNLSL 448  
 QY 376 L 376  
 Db 449 L 449

## RESULT 5

E95934  
 C:Species: *Sinorhizobium meliloti*  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: E95934  
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo  
 A:Reference number: A95842; MUID:21396508; PMID:11481431  
 A:Accession: E95934  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-307 <KUR>  
 A:Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:g15140626; GSPDB:GN00167  
 A:Experimental source: Strain 1021, megaplasmid pSymB  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 pel, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaune,  
 hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Welis, D.H.; Wong, K.; Yen, K.  
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: SMD21237  
 A:Genome: plasmid

Query Match 8.2%; Score 168; DB 2; Length 307;  
 Best Local Similarity 22.7%; Pred. No. 9.2e-07;  
 Matches 83; Conservative 50; Mismatches 131; Indels 102; Gaps 16;

QY 36 MKAQPERMAYLVLSWRSGSSFFVQGLFGQHPDYF---YLMERPAHVMWTFKOSTAMLMH 92  
 Db 1 MFSQPVATAYIAGYGRSSGTTLDIALGQHAANYGAGETISLTRHWRENEYCA----- 53  
 QY 93 AVRDILRAVELCOMSVFADAMEBGRPROSSLFQWENSRLCSAPACDIIPODETIIRAH 152  
 Db 54 -----CGNAIRDCSFMSSVRR-----FMSDG-----QDPGLMEYXC 84  
 QY 153 RLCSQQPEFVEVKACR-----SYSHVYLKEVRFN-IQS----- 186  
 Db 85 AL---QQKFEGLSMATLTLSGMLGKQFSLYIHTKRLTSAMOSCGRGROYVDSKLPGR 141  
 QY 187 LYPCLKPSLNLHIVLRDPRAVFRS-----RERTKGDLMTDSRTVMQHEQKAKED 240  
 Db 142 AMVAQAQPGIDMIVYHIVRGGRGAVMSLKGVERDKSG-----IQKEL 185  
 QY 241 QPIYVWVYVQSOLEIKTKTQSLPKAL-QERYLLVRYEDLARPAVQTSRMEFVGLFEL 299  
 Db 186 KPSVSVFTALRWSM-VLAVEVLSRKISGSKVRKVRVEEFASDPVAVMOOIGTFLELD-L 243  
 QY 300 PLOTWVHNITRGKMG-----DHAFTNARDALNVSOAMRWSLVEKVSRLQKACGDA 353  
 Db 244 SQTSTSEN---GEAMPGHGVAGNRLKMASTALKKDETRIRMPARQOVSPORLGGM 300  
 QY 354 MNLGY 359  
 Db 301 LRRGY 306  
 RESULT 6  
 G70729  
 hypothetical protein Rv2267c - *Mycobacterium tuberculosis* (strain H37Rv)

C:Species: *Mycobacterium tuberculosis*  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: G70729  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 ; Connor, R.; Davies, R.; Devlin, K.; Fielding, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
 A:Reference number: A70500; MUID:96295987; PMID:9634230  
 A:Accession: G70729  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-388 <COL>  
 A:Cross-references: GB:Z77163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e25507  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: Rv2267c

Query Match 5.7%; Score 116; DB 2; Length 388;  
 Best Local Similarity 19.3%; Pred. No. 0.037;  
 Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VLVSSWRSGSSFFVQGLF---GQH---PDVYLMERPAHVMWTFKOSTAMLMAYRDLIR 99  
 Db 83 IFVGHMRTGTLHLHLVVDNRHTGPGYECLEP--HFELD---TEW----- 125  
 QY 100 AVFLCOMSVFADAMEBGRPROSSLFQWENSRLCSAPACDIIPO-DETI-----PRAHC 152  
 Db 126 -----FABYVE-----FLVSKHRKMDNDLSLHHPQEDERFWCMQGLPSYL 167  
 QY 153 RLCSQQPEFVEVKACR---SYSHVYLKEVRFN-IQS-----PLKPSLNLH 200  
 Db 168 TIAFPRPQVEEYLDLEQVARELE--IKRLELRFQVYFRKRVILKPNHPSRI 225  
 QY 201 -----VHVRDPRAVFRSREKTKGDLMTDSRTVMQHEQKAKEDQPIY--VM 246  
 Db 226 KVLVEFPOAKRIHVRDPRVYVY-----PSTIHLKALRYHGLQDPFEDLD 273  
 QY 247 QVICSOLEIKTKTQSLPKALQ-ERYLLVRYEDLARPAVQTSRMEFVGLF---FLPH 301  
 Db 274 KAVSVTVVDLYKRLDGRILVPTFYEIRYEDLDLGDGQLRKLYQNLGLGDFECYLR 333  
 QY 302 LOTWVHNITRGKMGDHA-FHTNARDALNVSOAMRWSLVEKVSRLQKACDAMNLGY- 359  
 Db 334 LQY-----LADHADYKTN-----YQLTVEQRAIVDEHWGHIIDRYGD 373  
 QY 360 RH 361  
 Db 374 RH 375

## RESULT 7

I40485  
 surfactin synthetase component I - *Bacillus subtilis*  
 N:Alternate names: competence protein srfA; surfactin production protein srfA; surf  
 C:Species: *Bacillus subtilis*  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 03-Nov-2000  
 C:Accession: I40485; B69718; S3517; A37323; S46967; A43705; S34985  
 R:Cosminu, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van S  
 M.O. Microbiol. 8, 821-831, 1993  
 A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthe  
 A:Reference number: I40485; MUID:93360813; PMID:8355609  
 A:Accession: I40485  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-358 <RES>  
 A:Cross-references: EMBL:X70356; NID:9396480; PIDN:CAA49816.1; PID:g396481  
 A:Experimental source: Strain M168 derivative of Jn642  
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Calzavara, B.; Capuano, V.; Carter, N.M.;  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, K.D.; Farrell, C.; Ferrari,  
 Nature 390, 249-256, 1997



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:18:46 ; Search time 40 Seconds

(without alignments)  
927.697 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038

Sequence: 1 MLPPKMKLLLEVSQMAIL.....EQRLNLLDLSTWVPEQIR 386

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	30.6	484	2	JC7350 N-acetylglucosamin
2	606.5	29.8	486	2	JC7351 N-acetylglucosamin
3	588.5	28.9	484	2	JC7351 N-acetylglucosamin
4	500.5	24.6	458	2	JE0261 chondroitin 6-sulf
5	168	8.2	307	2	E55934 probable enzyme, C
6	116	5.7	388	2	G70729 hypothetical prote
7	95	4.7	358	2	140485 surfactin syntheta
8	92.5	4.5	2354	2	T13288 mei-41 protein - f
9	92	4.5	359	2	T16350 hypothetical prote
10	91.5	4.5	338	2	JE0196 hydroxysteroid sul
11	91	4.5	659	2	S11737 resistance protein
12	91	4.5	1847	2	E64477 replicational factor
13	89.5	4.4	346	2	E86319 probable flavonol
14	89.5	4.4	504	2	T07120 probable cytochrom
15	89.5	4.4	4273	2	C69679 polyketide synthas
16	89	4.4	524	1	A36205 mitochondrial proc
17	89	4.4	1140	2	G71335 probable transcrip
18	89	4.4	1968	1	P00093 genome polyprotein
19	88	4.3	359	2	T47187 hypothetical prote
20	88	4.3	466	2	E84132 aminopeptidase B13
21	87.5	4.3	380	2	T42755 tyrosylprotein sul
22	87.5	4.3	4063	2	T42993 probable spectrin
23	87.5	4.3	4101	2	T23630 hypothetical prote
24	87	4.3	2500	2	G88493 protein F57B9.2 [1
25	86.5	4.2	748	1	S08680 methylmalonyl-CoA
26	86.5	4.2	831	2	S26675 DNA-directed DNA p
27	86.5	4.2	2352	2	T06077 splicing factor PR
28	86	4.2	320	2	A40216 flavonol 4'-sulfot
29	86	4.2	823	2	S48986 probable protein k

30	86	4.2	869	2	A95558 probable protein k
31	85.5	4.2	344	2	H83928 hypothetical prote
32	85.5	4.2	764	1	S14113 1-phosphatidylinos
33	85	4.2	632	2	T18692 hypothetical prote
34	85	4.2	672	2	T34898 hypothetical prote
35	85	4.2	1237	2	T37529 hypothetical prote
36	85	4.2	2688	2	T49477 alpha-A-crystallin
37	84.5	4.1	217	2	J01358 CI protein - MscA
38	84.5	4.1	646	2	G86863 threonine-tRNA lig
39	84.5	4.1	1031	2	T43458 hypothetical prote
40	84.5	4.1	2493	2	S45734 probable membrane
41	84	4.1	390	2	AB2337 group 2 sigma 70-t
42	84	4.1	1189	2	T42726 guanine nucleotide
43	84	4.1	1217	2	T00270 hypothetical prote
44	84	4.1	1583	2	T14176 probable phosphati
45	83.5	4.1	518	2	E84536 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

JC7350 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)

C:Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 31-Dec-2000

C:Accession: JC7350

R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, R.; Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of

A:Reference number: JC7350; MUID:20374462; PMID:10913333

A:Accession: JC7350

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosyl

sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand

C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.6% Score 623; DB 2; Length 484;

Best Local Similarity 34.2% Pred. No. 1.56-45;

Matches 136; Conservative 65; Mismatches 151; Indels 46; Gaps 7;

QY	30	NISSLSMKAOPERMHVLSSWSGSSFFVQGLFGHPDVFILMEPAHVMWTFKOSTAM	89
DB	87	NLSAAGEAVTQEKQHYVHATMTRCSSLGLGFNFHPVFILYEPMMHLMQALYRGALES	146
QY	90	IMHAYRDLIRAVFLIDMSVFDAYMEGPRQ-----SLPQWENSALCSAPACD	139
DB	147	LOGALRDMIRSLFRCDFSVLRLYAQPGDGERAPDSANLTYAMLFRRMTNVICSPPLCP	206
QY	140	IIPDQ-----IIPRAHCHLCSQGFVEVEKACRSYSHVLEKEYRFFVLQSLYPLKDP	195
DB	207	AAPRRADVGLVEDKACESTCEPVSRLALEACRKYPPVVIKDFLDELGLVLELRDPG	266
QY	196	LNHLIHLVDRPRAVERSERETKGLMDS-----RIYV-----GOHE	233
DB	267	LNKLVQYLFDRPRAVHNSRLKSRQGLRESIQVLTFRQGDHFHVLIAHGVDARPGQA	326
QY	234	QKTKEDQPYV---VMQVIOQSULEIKTKIOSLPKALQEVYLLVREEDLARAQAQTSR	289
DB	327	RALPAPAPADFFLTSALVEICAWLRDLFTRGAPAWLRRYLRKREEDLWQFOQARR	386
QY	290	MEEFGLFELPHLOTWVINITRGKMG-DHAFHTNAPDALVSOAWRSLPEYKVSRLQK	348
DB	387	LLRFGLFTFLAALDAFANMTRGSAVGAIDRPFHLASARARAVHVMRRLSQEQVQVET	446
QY	349	ACGDAMLLGYRHVRSDEQQRNL-----LDLSTWV 361	
DB	447	ACAPAMRLAYPRSGDRDKRTVREGTEPLETKAMWV 484	

RESULT 2

CC during transplantation.  
 CC Note: The present sequence is also shown in sequence listing (page CC no: 56) but lacks four nucleotides at its 3' end.  
 XX  
 SQ Sequence 390 AA;  
 Query Match 49.5%; Score 1008; DB 22; Length 390;  
 Best local Similarity 52.3%; Pred. No. 3.7e-97;  
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;  
 QY 1 MLPRKMKLLFLVSOAAILALFFHMYSHNISTLSMKAPERMHVLYLSSMRSGSSFFVQ 60  
 DB 14 LLLAQTCLLFLITSRP-----GSSPAGGEDRHHVLYLSSMRSGSSFFLQ 59  
 QY 61 LFGQHPDVEFLMEPAHWMTFKOSTAMLMHAYRDLIRAVFLCDMSVFDAYMEPPRRQ 120  
 DB 60 LFSQHPDVEFLMEPAHWMTFLSOGSAATLHMAYRDLMSIFLCDMDVDAYM-POSRL 118  
 QY 121 SSLFOWNSRALCSAPACDIIPODEIIPRAHCRILCSQGFVEVEKACRSYSHVLYKEVR 180  
 DB 119 SAFENMATSRLCSPACSAFPREGTISKQDYCKTCTROPFSILAREKRSYSHVLYKEVR 178  
 QY 181 FENLQSLVPLKDPSSLNLHIVLRDPRAVRSRERTKGDLMDSRIYMGQHEQKLRKD 240  
 DB 179 FENLQVLYPLSDPALNLRIHVLVDRPRAVRSREAGPILARDNGIVLGIN-GKWEAD 237  
 QY 241 QPYVMOVICOSQLEIK--TIQSLKALOERYLVREYEDLARAPVAQTSRMEEVGLGF 298  
 DB 238 PHLRLREVCRSHVRIAEATLKP-PFLRGRYLVREYEDLARAPVAQTSRMEEVGLGF 296  
 QY 299 LPHLQTVHNTIRGKGMGD--HAFTNARDALNVSOAMRMSLPEYKYSRIQKACGDAMNL 356  
 DB 297 TPQLEAMINHTHSGIGKRIEAFHTSSRNARNVSOAMRHALPFTKILRVQEVACAGALQ 356  
 QY 357 LGRHVRSSEQQRNLHDL-----STWTYPE 383  
 DB 357 LGRPVYSADQQRDLTLDVLPGRPDHFSWASPD 390  
 RESULT 15  
 ABB81556 standard; Protein: 390 AA.  
 ID ABB81556  
 AC ABB81556;  
 DT 05-SEP-2002 (first entry)  
 DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.  
 KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST; corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD; ophthalmological.  
 OS Homo sapiens.  
 XX US2002061562-A1.  
 XX 23-MAY-2002.  
 XX 09-AUG-2001; 2001US-0927602.  
 XX 11-AUG-2000; 2000US-325737P.  
 XX (FUKU//) FUKUDA M N.  
 XX (AKAM//) AKAMA T O.  
 XX Fukuuda MN, Akama TO;  
 XX WPI: 2002-507643/54.  
 XX New nucleic acid encoding corneal  
 XX N-acetylglucosamine-6-sulfotransferase, useful for treatment,  
 XX monitoring and diagnosis of macular corneal dystrophy

XX Example 5; Fig 2A-B; 69pp; English.  
 PS  
 CC The present invention describes human corneal  
 CC N-acetylglucosamine-6-sulfotransferase (1), which is able to catalyze  
 CC sulfation of keratan sulfate (KS). Also described is a method for  
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),  
 CC and detecting susceptibility to MCD. (1) is located to chromosome 16q22,  
 CC and has ophthalmological activity. (1) can be used to treat or prevent  
 CC macular corneal dystrophy types I or II. (1) makes possible treatment  
 CC of MCD without requiring keratoplasty or keratectomy. The present  
 CC sequence represents human intestinal N-acetylglucosamine-6-  
 CC sulfotransferase, which is given in comparison with (1) in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 390 AA;  
 Query Match 49.5%; Score 1008; DB 22; Length 390;  
 Best local Similarity 52.3%; Pred. No. 3.7e-97;  
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;  
 QY 1 MLPRKMKLLFLVSOAAILALFFHMYSHNISTLSMKAPERMHVLYLSSMRSGSSFFVQ 60  
 DB 14 LLLAQTCLLFLITSRP-----GSSPAGGEDRHHVLYLSSMRSGSSFFLQ 59  
 QY 61 LFGQHPDVEFLMEPAHWMTFKOSTAMLMHAYRDLIRAVFLCDMSVFDAYMEPPRRQ 120  
 DB 60 LFSQHPDVEFLMEPAHWMTFLSOGSAATLHMAYRDLMSIFLCDMDVDAYM-POSRL 118  
 QY 121 SSLFOWNSRALCSAPACDIIPODEIIPRAHCRILCSQGFVEVEKACRSYSHVLYKEVR 180  
 DB 119 SAFENMATSRLCSPACSAFPREGTISKQDYCKTCTROPFSILAREKRSYSHVLYKEVR 178  
 QY 181 FENLQSLVPLKDPSSLNLHIVLRDPRAVRSRERTKGDLMDSRIYMGQHEQKLRKD 240  
 DB 179 FENLQVLYPLSDPALNLRIHVLVDRPRAVRSREAGPILARDNGIVLGIN-GKWEAD 237  
 QY 241 QPYVMOVICOSQLEIK--TIQSLKALOERYLVREYEDLARAPVAQTSRMEEVGLGF 298  
 DB 238 PHLRLREVCRSHVRIAEATLKP-PFLRGRYLVREYEDLARAPVAQTSRMEEVGLGF 296  
 QY 299 LPHLQTVHNTIRGKGMGD--HAFTNARDALNVSOAMRMSLPEYKYSRIQKACGDAMNL 356  
 DB 297 TPQLEAMINHTHSGIGKRIEAFHTSSRNARNVSOAMRHALPFTKILRVQEVACAGALQ 356  
 QY 357 LGRHVRSSEQQRNLHDL-----STWTYPE 383  
 DB 357 LGRPVYSADQQRDLTLDVLPGRPDHFSWASPD 390  
 Search completed: January 11, 2003, 01:18:39  
 Job time : 76 secs

**This Page Blank (uspto)**

```

FT Misc-difference 371
FT /label= Ala, Thr, Ser
FT Misc-difference 380
FT /label= Leu, Pro, Met
FT Misc-difference 382
FT /label= Gly, His, Ser
FT Misc-difference 384
FT /label= Thr, Ser, Lys
FT Misc-difference 390
FT /label= Ala, Glu
FT Misc-difference 391
FT /label= Ser, Lys
FT Misc-difference 392
FT /label= His, Glu
FT Misc-difference 394
FT /label= Arg, Glu
FT Misc-difference 395
FT /label= Asn, Ser
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001, 2001US-0927602.
XX
XX 11-AUG-2000, 2000US-325773P.
XX
XX (FUKU/) FUKUDA M. N.
XX (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI: 2002-507643/54.
XX
XX New nucleic acid encoding corneal
XX N-acetylglicosamine-6-sulfotransferase, useful for treatment,
XX monitoring and diagnosis of macular corneal dystrophy -
XX
XX Example 5; Fig 2A-B; 69pp: English.
XX
XX The present invention describes human corneal
XX N-acetylglicosamine-6-sulfotransferase (I), which is able to catalyze
XX sulfation of keratan sulfate (KS). Also described is a method for
XX monitoring the effect of treatments for macular corneal dystrophy (MCD),
XX and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
XX and has ophthalmological activity. (I) can be used to treat or prevent
XX macular corneal dystrophy types I or II. (I) makes possible treatment
XX of MCD without requiring keratoplasty or keratectomy. The present
XX sequence represents a consensus N-acetylglicosamine-6-sulfotransferase
XX which is given in the exemplification of the present invention.
XX
XX Sequence 395 AA:
XX
XX Query Match 49.7%; Score 1013.5; DB 23; Length 395;
XX Best Local Similarity 55.9%; Pred. No. 9,8e-98;
XX Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;
XX
XX 2 LIPKMKLLFLVYSOMAILALFFHMYSHNITSLSMKAPRHHVLYLSMRSSGSSEVQGL 61
XX | : : : : : |
XX 14 LLLAQTKLLFLVSRP-----GSSSPAGEKRVHLYLVSMRSGSSEVQGL 59
XX | : : : : : |
XX 62 FGQHPDYFLMEPAWMMTFKOSTAMLMHAYVDLIRAVFLCDMSVFDAYMEPRRQS 121
XX | : : : : : |
XX 60 FQSPDVFYLMPEAWHWTLSQSSAATLHMAVADLKRVSFLCDMDVFDAYL-PWRNLS 118
XX | : : : : : |
XX 122 SLFQWMSRALCSPACDITIPDEIIPRACRLICSQPFEEVERKACRSYSHVLKVF 181
XX | : : : : : |
XX 119 DLFQWMSRALCSPACSAFPRKXISEYCKPLCARQPPXIAREKRSYSHVLKVF 178
XX | : : : : : |
XX 182 FNLQSLYPLKDPCLNHIYLVDPRAVPRSRRTKGDLMIDSRIVYMGHEOKLKEDQ 241
XX | : : : : : |
XX 179 FNLQSLYPLKDPCLNHIYLVDPRAVPRSRRTKGDLMIDSRIVYMGHEOKLKEDQ 237

```

```

QY 242 PYVWQVICOSOLEIYK--TIQSLPKALQERILVRYEDLARAAYAGTSRMTEFVGLEFL 299
QY | : : : : : |
DB 238 XLRVAREVCRSHVRIAEATLKP--PPFLGRRIYRFEDLAREPLAEIRAYAEFTGKLT 296
QY 300 PHLQFWANITPRKGMG--DHAEHTNARDALNVSQAWMSLPEYEVSPLOKACDAMNL 357
DB | : : : : : |
DB 297 POLEMINHNTHTSGPGARREAFKTSRNALNVSQAWHMLPFAKIRVQGLCAGALQL 356
QY 358 GYRVRSEDEQRNLLDL 376
DB | : : : : : |
DB 357 GYRVPYSEDEQRDLXLDLV 375

RESULT 14
AAV72639
ID AAV72639 standard; Protein: 390 AA.
XX
XX AAV72639;
XX
XX 02-MAY-2001 (first entry)
XX
XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX
XX Human glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polypositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
XX
XX Homo sapiens.
XX
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US19741.
XX
XX 20-JUL-1999; 99US-0144694.
XX
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX
XX WPI: 2001-138471/14.
XX
XX N-PSDB; AAD02697, AAD02698, AAD02699.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications -
XX
XX Claim 3; Fig 1; 128pp: English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha). GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polypositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

```

QY 298 FLPHLQTVHNITRGKMG--DHAFTNARDALVSOAMRWSLPEYKVSRLQKAGDAMN 355  
 Db 295 LTPQLQTVHNITRGSGPGARREAFKTSRDALSVQAMRHLPFAKIRVOELCGALQ 354  
 QY 356 LGGYRVHSEJOEORNLIDL 376  
 Db 355 LGGYRVHSELEQORDSLIDL 375

## RESULT 12

AAU11275  
 ID AAU11275 standard; Protein: 395 AA.

AC AAU11275;

DT 12-MAR-2002 (first entry)

XX Marine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX Mouse: beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;  
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;  
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;  
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;  
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;  
 KW antileukematory; antipsoriatic; antidiabetic; dermatological;  
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.

XX Mus musculus.

XX WC020185177-A1.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001MO-US15452.

XX 11-MAY-2000; 2000OOS-0569320.

XX (BURN-) BURNHAM INST.

XX Fukuda M, Yeh J, Hiraoka N.

XX WPI; 2002-075226/10.

XX N-PSDB; AAS16948.

PT New enzyme, useful for modifying acceptor molecule, comprises an  
 PT isolated L-selectin sulfotransferase-2 that directs expression of  
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or  
 PT intestinal GlcNAc 6-sulfotransferase

PS Claim 28; Fig 10; 98pp; English.

XX The present invention provides a method of modifying an acceptor molecule  
 CC by contacting the acceptor with an isolated  
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active  
 CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The  
 CC invention also provides a method of treating or preventing an  
 CC L-selectin-mediated condition by reducing the expression or activity of a  
 CC beta1,3GNT that directs expression of a MECA-79 antigen. This can be done  
 CC by administering to the subject an oligosaccharide L-selectin antagonist  
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by  
 CC administering antibody material that specifically binds beta1,3GNT,  
 CC and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin  
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.  
 CC Alternatively, the expression or activity of LST-2 or its active  
 CC fragment can be reduced in combination with reducing the expression or  
 CC activity of beta1,3GNT. The method is useful for treating L-selectin  
 CC mediated conditions such as Crohn's disease and ulcerative colitis;  
 CC inflammatory disorders of the skin such as allergic contact dermatitis,  
 CC psoriasis and lichen planus; lymphomas, chronic pneumonia, delayed-type  
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
 CC sequence represents mouse I-GlcNAc6ST.

XX Sequence 395 AA;

Query Match 49.9%; Score 1017.5; DB 23; Length 395;  
 Best Local Similarity 56.4%; Pred. No. 3.7e-98;  
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLPK--KKKLLFLVSQAIALLFFHMTSHNLSLTKMAQEPKMAHYLVSSMRGSSFFV 58  
 Db 1 MRLPRFSSTVMISLIMVQIGLVE--LVSQVPS--SPAGLGERVHYLVLSMRGSSFFV 56  
 QY 59 GOLFQGPVPEFLMEBPAMVHMTFKOSTAMLMMAVRDLIRAFVLCDSMVPAVKEPGR 118  
 Db 57 GOLFQGPVPEFLMEBPAMVHMTFKOSTAMLMMAVRDLIRAFVLCDSMVPAVKEPGR 115  
 QY 119 ROSSLFQWENSRLACAPACDIIIPDEILPRAHCRILCSQPFVEYKACRSYSHVLKE 178  
 Db 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRFFGIAQECSSYSHVLKE 175  
 QY 179 VRFNFIQSLYPLIKDLSLHVIHLVDRDRAVERSEKPGDMLDSRTVMQCHQKTK 238  
 Db 176 VRFNFIQSLYPLIKDLSLHVIHLVDRDRAVERSEKPGDMLDSRTVMQCHQKTK 234  
 QY 239 EDQPIYVQVICOQSOLEIK-TIQSLPKALOERYLLVRYEDLARAPVAQTSMEYFVGL 297  
 Db 235 ADPRLVAVNEGVCRSHVRIEALHKPPPLQDRYLRVREYEDLARAPVAQTSMEYFVGL 294  
 QY 298 FLPHLQTVHNITRGKMG--DHAFTNARDALVSOAMRWSLPEYKVSRLQKAGDAMN 355  
 Db 295 LTPQLQTVHNITRGSGPGARREAFKTSRDALSVQAMRHLPFAKIRVOELCGALQ 354  
 QY 356 LGGYRVHSEJOEORNLIDL 376  
 Db 355 LGGYRVHSELEQORDSLIDL 375

## RESULT 13

ABB81555  
 ID ABB81555 standard; Protein: 395 AA.

XX ABB81555;

XX 05-SEP-2002 (first entry)

XX Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO.3.

KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GlcNAc6ST;  
 KW cornell; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
 KW ophthalmological.

XX Homo sapiens.

XX Mus musculus.

XX Synthetic.

FT Key Location/Qualifiers  
 FT Misc-difference 10 /Label= Ala, Thr, Val  
 FT Misc-difference 13 /Label= Ala, Val, Ser  
 FT Misc-difference 20 /Label= Phe, Cys, Gly  
 FT Misc-difference 39 /Label= Ala, Asp, Glu  
 FT Misc-difference 96 /Label= Val, Met, Ile  
 FT Misc-difference 142 /Label= Ala, Thr, Asn  
 FT Misc-difference 147 /Label= Ala, Asp, Glu  
 FT Misc-difference 159 /Label= Thr, Ser, Gly  
 FT Misc-difference 238 /Label= Gly, His, Arg  
 FT Misc-difference 294 /Label= Ser, Thr, Gly



CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,  
 CC asthma, autoimmune hemolytic anaemia, contact dermatitis, Crohn's  
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's  
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple  
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus  
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,  
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic  
 CC infections; cell proliferative disorders such as actinic keratosis,  
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,  
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,  
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus  
 CC and pituitary resulting from lesions such as primary brain tumours,  
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular  
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,  
 CC glaucoma, pigmentosa, metabolic disorders such as Addison's disease,  
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,  
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,  
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such  
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,  
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's  
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental  
 CC disorders. The present sequence is human DME-5 protein.

XX Sequence 395 AA:

Query Match 50.0%; Score 1019; DB 23; Length 395;

Best Local Similarity 54.5%; Pred. No. 2.6e-98;

Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LIPKMKLLFLVSQMAILLFFHMYSHNISLSMKAOBERNHVILSSWRSGSSFFVGL 61  
 DB 14 LLAAQTFLLFLFVSRP-----GPSSPAGCBARVAVLTLSSWRSGSSFFVGL 59  
 QY 62 FQGHDPVFLMEPAHVMWMTFKOSTAMKLMAYRDLIRAVFLCDMSVPAAYMEPCRR 121  
 DB 60 FQGHDPVFLMEPAHVMWMTLISQSGATLHMAYRDLIRAVFLCDMSVPAAYL-PWRN 118  
 QY 122 SLFQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQGFVEVAKRSYSHVILKEVRF 181  
 DB 119 DLFQWNSRALCSPPACSAFPGATISSEAVCKFLCAROSFTLAREACRSYSHVILKEVRF 178  
 QY 182 FNIQSLYPLIKDPSLNHTIHLVDRPRAVRSRERTKGLMISRTVMQHEKKLKEQ 241  
 DB 179 FNIQSLYPLIKDPSLNHTIHLVDRPRAVRSRERTKGLMISRTVMQHEKKLKEQ 237  
 QY 242 PYVWQVTCOSQLEIYK--TIQSLPKALOERYLLVREDIARAPVAQTSRMVEFVLEPL 299  
 DB 238 GLRVREVCRSYRIAEATLKP--PPFLRGRYLVREDIARAPVAQTSRMVEFVLEPL 296  
 QY 300 PHIQTVNHTTRKGMG--DHAFTNARQALNSQAMRNSLPREKVSRLQKAGCDAMNL 357  
 DB 297 POEAMVHNHTHSGPARREAFKISSRNALNSQAMRHALLPFAKIRVOELCAGALQL 356  
 QY 358 GYRHVSEDFQORLLDL-----STWTVPEQIH 386  
 DB 357 GYRPVYSEDFQORLLDLVLRGLNFTMASSTASH 392

RESULT 11  
 AAAY72638  
 ID AAAY72638 standard; Protein; 395 AA.

XX AAAY72638;

XX 02-MAY-2001 (first entry)

XX Mouse glycosyl sulfotransferase-4 (GST-4).

KM Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive;  
 KM therapy; selection binding inhibitor; gene therapy; inflammation;  
 KM systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;  
 KM polyarteritis nodosa; polyomyelitis; systemic sclerosis; dermatitis;  
 KM glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;

KM Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;  
 KM demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;  
 KM myocarditis; adult respiratory distress syndrome; eczema; psoriasis;  
 KM asthma; hypersensitivity; rheumatic fever; tissue rejection;  
 KM chromosome 8p1.

XX Mus musculus.

OS Mus musculus.

XX W0200106015-71.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WD-US19741.

XX 20-JUL-1999; 99US-0144694.

XX 13-JUL-2000; 2000US-0538282.

XX (REGC ) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerlich S;

XX WPL: 2001-138471/14.

XX N-PSDB; AAD02696.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for  
 PT diagnostic and therapeutic agent screening applications -  
 PS Claim 3; Fig 2; 128pp; English.

XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4  
 CC gene is found on chromosome 8p1.

CC GST is a type 2 membrane protein useful for inhibiting a binding event  
 CC between a selectin and a selectin ligand, which comprises contacting the  
 CC selectin with a non-sulphated selectin ligand. GST and a small molecular  
 CC agent that inhibits the sulphation activity of GST. GST is also useful  
 CC in inhibiting a selectin mediated binding event. GST is useful in gene  
 CC therapy to treat disorders such as acute or chronic inflammation,  
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis  
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,  
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's  
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, perniciosis  
 CC anemia, demyelinating diseases, cirrhosis, ulcerative colitis,  
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress  
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,  
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection  
 CC during transplantation.

XX Sequence 395 AA:

Query Match 49.9%; Score 1017.5; DB 22; Length 395;  
 Best Local Similarity 56.4%; Pred. No. 3.7e-98;  
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLPK--KKKLLFLVSQMAILLFFHMYSHNISLSMKAOBERNHVILSSWRSGSSFFV 58  
 DB 1 MLPPFSSIVMLSLMGCGILVF--LVSRQVPS-SPAGLGERHVLTLSSWRSGSSFFV 56  
 QY 59 GOLFGHDPVFLMEPAHVMWMTFKOSTAMKLMAYRDLIRAVFLCDMSVDAAYMEPCRR 118  
 DB 57 GOLFGHDPVFLMEPAHVMWMTLISQSGAPALHMAYRDLIRAVFLCDMSVDAAYL-PWR 115  
 QY 119 ROSSLFQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQGFVEVAKRSYSHVILKEVRF 178  
 DB 116 NISDLFQWNSRALCSPPVCEAFAGNINISSEVCKFLCATNRFGLAQACSSYSHVILKE 175  
 QY 179 VRFENQSLYPLIKDPSLNHTIHLVDRPRAVRSRERTKGLMISRTVMQHEKKLKEQ 238  
 DB 176 VRFENQSLYPLIKDPSLNHTIHLVDRPRAVRSRERTKGLMISRTVMQHEKKLKEQ 234  
 QY 239 EQQPYVWQVTCOSQLEIYK--TIQSLPKALOERYLLVREDIARAPVAQTSRMVEFVLE 297  
 DB 235 ADPRLLRVNVECRSHVRIAEALHKPPFLDRLRYLVREDIARAPVAQTSRMVEFVLE 294

```

XX AC ABB81554;
XX DT 05-SEP-2002 (first entry)
XX DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX KW Humal: N-acetylglucosamine-6-sulfotransferase; enzyme: GlcNAc6ST;
XX KW Corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX KW ophthalmological; chromosome 16q22.
XX OS Homo sapiens.
XX PN US2002061562-A1.
XX PD 23-MAY-2002.
XX PF 09-AUG-2001: 2001US-0927602.
XX PR 11-AUG-2000: 2000US-325773P.
XX PA (FUKU/) FUKUDA M N.
XX PA (AKAM/) AKAMA T O.
XX PI Fukuda MN, Akama TO;
XX DR WPI: 2002-507643/54.
XX DR N-PSDB: ABB89506.
XX PT New nucleic acid encoding corneal
XX PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
XX PT monitoring and diagnosis of macular corneal dystrophy
XX PS Claim 13: Fig 1A-D: 69pp: English.

CC CC The present sequence represents human corneal
CC CC N-acetylglucosamine-6-sulfotransferase (1), which is able to catalyze
CC CC sulfation of keratan sulfate (KS). Also described is a method for
CC CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC CC and detecting susceptibility to MCD. (1) is located to treat or prevent
CC CC macular corneal dystrophy types I or II. (1) makes possible treatment
CC CC of MCD without requiring keratoplasty or keratectomy.
XX CC
XX SQ Sequence 395 AA;
Query Match 50.0%; Score 1019; DB 23; Length 395;
Best Local Similarity 54.5%; Pred. No. 2,6e+98;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LERKMLLFVLSQMAIALPFFHMYSHNITSLSKRQPERHVLVSSMRSGSFFVQGL 61
DB 14 LLAQGLFLFLVSRP-----GPSPPAGGEARVHVLVSSMRSGSFFVQGL 59
QY 62 FGOHPDYFYIMEBAHVWMTFKQSTAMLMHAYVDLIRAVFLCDMSVFPAYMEGPRROS 121
DB 60 FNOHPDYFYIMEBAHVWMTLSSQSAATLHMAVRDLVRSVFLCDMDVFPAYL-PMRRNLS 118
QY 122 SLFQWENSRAICSPACDIIPODEIIPRAHCRLLCSQPFEEVEKCRSSHYLVKEVFE 181
DB 119 DLFQWANSRAICSPACSAFPRGAISSAVCKPLCAROSTTLARECRSHVLYKEVFE 178
QY 182 FNIQSLYPLIKDPSLNHIVLVDRPRAVRSREKRGKGLMDISRIVNGOHEQKLEKQD 241
DB 179 FNIQSLYPLIKDPSLNHIVLVDRPRAVRSREKRGKGLMDISRIVNGOHEQKLEKQD 241
QY 242 PYYVMOVICQSOLEIK--TIQSLPKALOERYTLVREPLARPVQOTSRMVEFVLEFL 299
DB 238 GLRVRFVCSHVRIADATLKP-PPPLKORYFLVRFEDLARPELALTAFTGLSLT 296
QY 300 PQLQIWNHTTRKSGMG--DHAFTNARDALNVSQAMRMSLPYEXKSRLOKACGDAMNLL 357
DB 297 PQLQIWNHTTRKSGMG--DHAFTNARDALNVSQAMRMSLPYEXKSRLOKACGDAMNLL 357

```

```

QY 358 GYRHWSEDEQRNLLDILL-----STWVPEQHT 396
DB 357 GYRHWSEDEQRNLLDILLPRLGNGFTWASSSTASH 392
RESULT 10
AAE15438
ID AAE15438 standard; Protein: 395 AA.
AC AAE15438;
AC AAE15438;
DT 12-MAR-2002 (first entry)
DE Human drug metabolizing enzyme (DME)-5.
XX XX
XX XX Human: drug metabolizing enzyme; gene therapy; autoimmune disorder;
XX XX inflammatory disorder; acquired immune deficiency syndrome; infection;
XX KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
XX KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
XX KW cancer; endocrine disorder; hypohalimus disorder; pituitary disorder;
XX KW gastrointestinal disorder; metabolic disorder; developmental disorder;
XX KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
XX KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
XX KW DME-5.
XX OS Homo sapiens.
XX FH Key
XX FT Peptide
XX FT /label= Signal_peptide
XX FT /note= "Human mature DME-5 protein"
XX FT Peptide
XX FT /label= Signal_peptide
XX FT /note= "Human mature DME-5 protein"
XX PD WO200179468-A2.
XX PF 12-APR-2001: 2001WO-US11869.
XX PR 13-APR-2000: 2000US-197590P.
XX PR 19-APR-2000: 2000US-198403P.
XX PR 28-APR-2000: 2000US-200185P.
XX PR 05-MAY-2000: 2000US-202234P.
XX PR 11-MAY-2000: 2000US-203509P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
XX PI Yuo MG, Yue H, Tang YT, Patterson C, Das D, Sanjapala MS;
XX PI Gaothi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Gilfiln JA;
XX PI Au-Young J;
XX DR WPI: 2002-066363/09.
XX DR N-PSDB: AAD24670.
XX PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
XX PT useful for diagnosing, treating, or preventing disorders associated
XX PT with aberrant expression of DME such as allergy, anemia, asthma,
XX PT infertility
XX PS Claim 1a; Page 131-132; 143pp: English.

CC CC The invention relates to human drug metabolizing enzymes referred as
CC CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
CC CC of the invention are useful for assessing toxicity of test compounds
CC CC and in gene therapy. Sequences of the invention are useful in the
CC CC diagnosis, prevention and treatment of autoimmune/inflammatory
CC CC disorders such as acquired immune deficiency syndrome (AIDS), adult

```



RESULT 6  
ID AAB41947 standard; protein; 418 AA.  
AC AAB41947;  
XX 08-FEB-2001 (first entry)  
DE Human ORFX ORF111 polypeptide sequence SEQ ID NO:3422.  
XX  
XX Human: open reading frame: ORFX; detection; cytosolic; hepatotropic;  
KW valineary; antipsoptic; antiparkinsonian; noctropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinephritic;  
KW antiviral; antibacterial; antifungal; antinephritic; antinephritic;  
KW antinephritic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
XX Homo sapiens.  
XX  
XX MO200058473-A2.  
XX  
XX 05-OCT-2000.  
XX  
XX 31-MAR-2000; 2000WO-US08621.  
XX  
XX 31-MAR-1999; 9905-0127607.  
XX  
XX 02-APR-1999; 9905-0127636.  
XX  
XX 05-APR-1999; 9905-0127728.  
XX  
XX 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shinkets RA, Leach M;  
XX  
XX WPI: 2000-602362/57.  
XX  
XX N-PSDB; AAC76156.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX  
XX Claim 11: Page 2599-2600; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; valineary;  
CC antipsoptic; antiparkinsonian; noctropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinephritic; antibacterial; antiviral; antifungal; antinephritic;  
CC antinephritic; and antinephritic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX

SEQ Sequence 418 AA:  
Query Match 59.2%; Score 1205.5; DB 21; Length 418;  
Best Local Similarity 61.2%; Pred. No. 6,7e-118;  
Matches 241; Conservative 47; Mismatches 79; Indels 27; Gaps 6;  
QY 1 MLPPKMKLLFLFVSOMATLALFFHYSHNIISSLSMKAOPERMHVLVLSWRSSSEVQ 60  
DB 41 LLLAQTCCLLFLTISRP-----GPPSPAGEDRVHVLVLSWRSSSEVQ 86  
QY 61 LFGQHADVYVLMPPANHYMTKQSTAMMLHMAVBDLRAVFLCMSVPDAMERGPRRO 120  
DB 87 LFSQHPDVEYVLEPAHWVTLTSSQSAATLHMAVBDLMSITFLCDMDVFDAMEGPRRO 146  
QY 121 SSLFQWENSRALCSAPACDIIPQDEIIPRACRCLLSCQPEFEVAKRSYSHVVLKEVR 180  
DB 147 SSLFQWENSRALCSAPACDIIPQDEIIPRACRCLLSCQPEFEVAKRSYSHVVLKEVR 206  
QY 181 FENLQSLYLLDPSLNLIHIVLRDPRAVFSSRERTKGLMIDRIYVGGHEQLKED 240  
DB 207 FENLQSLYLLDPSLNLIHIVLRDPRAVFSSRERTKGLMIDRIYVGGHEQLKED 265  
QY 241 QPYVWQVICSQLEIYK--TQSLEPKALQERYLLVRYEDLARAPVQTSRMYEFVGLF 298  
DB 266 PHLRRLREYCRSHVRIAEATILKP--PPLRGRIYLRVEDLARAPVQTSRMYEFVGLF 324  
QY 299 LPHLQVWVHNITRGKQMD--HAFTNAPDALVNSQAMRSLPYEKVSLQACGDAMNL 356  
DB 325 TPLEAMINITHGSGICKEPIEAFHTSSRNARNVSOAMRHALPFTKILRVDCAGALQ 384  
QY 357 LGRHYRSFQEOBNLLDEL-----STWTYPE 383  
DB 385 LGRHYRSFQEOBNLLDEL-----STWTYPE 418  
RESULT 7  
ID ABB81557 standard; protein; 418 AA.  
XX  
XX ABB81557;  
XX  
XX 05-SEP-2002 (first entry)  
XX  
XX  
XX Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.  
DE  
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNA65T;  
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;  
KW ophthalmological.  
XX  
XX Mus musculus.  
XX  
XX US2002061562-A1.  
XX  
XX 23-MAY-2002.  
XX  
XX 09-AUG-2001; 2001US-0927602.  
XX  
XX 11-AUG-2000; 2000US-325773P.  
XX  
XX (FUKU/) FUKUDA M N.  
XX (AKAW/) AKAWA T O.  
XX  
XX Fukuda MN, Akawa TO;  
XX  
XX WPI: 2002-507643/54.  
XX  
XX New nucleic acid encoding corneal  
PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,  
PT monitoring and diagnosis of macular corneal dystrophy -  
XX  
XX Example 5; Page 24-25; 69pp; English.  
XX  
XX The present invention describes human corneal

PA (INCY-) INCYTE PHARM INC.  
 XX  
 XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;  
 PI Hillman JL, Azimzai Y;  
 XX  
 DR WPI: 2000-256996/22.  
 DR N-PSDB: AA294211.  
 XX  
 PT Human transferase proteins useful for preventing, diagnosing and  
 PT treating cancers and developmental, gastrointestinal, genetic,  
 PT immunological, neurological, reproductive and smooth muscle disorders  
 XX  
 PS Claim 1, Page 90-91, 113pp: English.  
 XX  
 CC The present sequence is that of human transferase TRNSFS-11, 1 of  
 CC 15 claimed novel human transferase proteins of the invention (see  
 CC AA79209-23). The sequence was deduced from a cDNA clone (see  
 CC AA294211) isolated from a gallbladder library. It shows homology to  
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is  
 CC expressed in dermatologic and gastrointestinal tissues,  
 CC especially those associated with inflammation and cell  
 CC proliferation. The new human transferases and polynucleotides can  
 CC be used in the diagnosis, prevention and treatment of cancer,  
 CC developmental disorders, gastrointestinal disorders, genetic  
 CC disorders, immunological disorders, neurological disorders,  
 CC reproductive disorders, and smooth muscle disorders. The  
 CC polypeptides can also be used to raise antibodies, and to screen  
 CC for agonists and antagonists of transferase activity.  
 CC  
 SQ Sequence 386 AA:  
 Query Match 95.0%; Score 1936; DB 21; Length 386;  
 Best Local Similarity 95.6%; Pred. No. 8.2e-195;  
 Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MLPPKMKLLFLVSGMALALFFHMYSHNISLSMKAQPERHNVILSSWRSGSSFFVQ 60  
 DB 1 MLPPKMKLLFLVSGMALALFFHMYSHNISLSMKAQPERHNVILSSWRSGSSFFVQ 60  
 QY 61 LEFGHPDVFYLMPEPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPPRRQ 120  
 DB 61 LEFGHPDVFYLMPEPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPPRRQ 120  
 QY 121 SSLEFQWENSALCSAPACDIIPDEIIPRAHCRLLCSQCFEVEVEKACRSYSHVLEKVR 180  
 DB 121 SSLEFQWENSALCSAPACDIIPDEIIPRAHCRLLCSQCFEVEVEKACRSYSHVLEKVR 180  
 QY 181 FENLQSLYPLKDPNLNHLVHVRDRAVFRSREBRTKGLMDSRIYMGQHEOKLKED 240  
 DB 181 FENLQSLYPLKDPNLNHLVHVRDRAVFRSREBRTKGLMDSRIYMGQHEOKLKED 240  
 QY 241 QPYVWVGTOSQOLEIKTIOSLPKALOERYLLVRYEDLARAPVAQTSRMVEFVGLFLP 300  
 DB 241 QPYVWVGTOSQOLEIKTIOSLPKALOERYLLVRYEDLARAPVAQTSRMVEFVGLFLP 300  
 QY 301 HLQTWVHNITRGKMGDAFHTNARDALNVSAWRMSLPEYKVSRLQACGDAMNLLGYR 360  
 DB 301 HLQTWVHNITRGKMGDAFHTNARDALNVSAWRMSLPEYKVSRLQACGDAMNLLGYR 360  
 QY 361 HVRSQEQQRNLDDLSTWTVEQI 386  
 DB 361 HVRSQEQQRNLDDLSTWTVEQI 386  
 RESULT 5  
 AA39919  
 ID AA39919 standard; Protein: 388 AA.  
 XX  
 AC AA39919;  
 XX  
 DT 08-DEC-1999 (first entry)  
 XX  
 DE Mouse glycosyl sulfotransferase-3 protein sequence.

XX  
 KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
 KW selectin binding interaction; inflammation; lymphocyte homing; mouse;  
 KW secondary lymph organ.  
 XX  
 OS Mus sp.  
 XX  
 PN W09949018-A1.  
 XX  
 PD 30-SEP-1999.  
 XX  
 PE 26-FEB-1999; 99WO-US04316.  
 XX  
 PR 20-MAR-1998; 98US-0045284.  
 XX  
 PR 12-NOV-1998; 98US-0190911.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (SYNT ) SYNTAX USA INC.  
 XX  
 PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;  
 XX  
 DR WPI, 1999-580442/49.  
 DR N-PSDB: AA220793.  
 XX  
 PT Human and murine glycosyl sulfotransferase 3 and related  
 PT polynucleotides  
 XX  
 PS Claim 2, Fig 4: 59pp: English.  
 XX  
 CC This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of  
 CC the invention. The nucleic acid sequences, probes and primers derived  
 CC from these, proteins and antibodies are useful in detecting homologues.  
 CC The sequences, antibodies and methods are useful in the diagnosis and  
 CC treatment of diseases associated with selectin binding interactions,  
 CC including conditions associated with or resulting from the homing of  
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
 CC to secondary lymph organs.  
 CC  
 SQ Sequence 388 AA:  
 Query Match 74.3%; Score 1513.5; DB 20; Length 388;  
 Best Local Similarity 72.8%; Pred. No. 2.4e-150;  
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;  
 QY 1 MLPPKMKLLFLVSGMALALFFHMYSHNISLSMKAQPER-MHVLVSSWRSGSSFFVQ 59  
 DB 1 MLPPKMKLLFLVSGMALALFFHMYSHNISLSMKAQPER-MHVLVSSWRSGSSFFVQ 59  
 QY 60 QLEFGHPDVFYLMPEPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPPRRQ 119  
 DB 60 QLEFGHPDVFYLMPEPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPPRRQ 119  
 QY 119 QSSLEFQWENSALCSAPACDIIPDEIIPRAHCRLLCSQCFEVEVEKACRSYSHVLEKVR 179  
 DB 119 QSSLEFQWENSALCSAPACDIIPDEIIPRAHCRLLCSQCFEVEVEKACRSYSHVLEKVR 179  
 QY 179 FENLQSLYPLKDPNLNHLVHVRDRAVFRSREBRTKGLMDSRIYMGQHEOKLKED 238  
 DB 179 FENLQSLYPLKDPNLNHLVHVRDRAVFRSREBRTKGLMDSRIYMGQHEOKLKED 238  
 QY 239 QPYVWVGTOSQOLEIKTIOSLPKALOERYLLVRYEDLARAPVAQTSRMVEFVGLFLP 299  
 DB 239 QPYVWVGTOSQOLEIKTIOSLPKALOERYLLVRYEDLARAPVAQTSRMVEFVGLFLP 299  
 QY 299 HLQTWVHNITRGKMGDAFHTNARDALNVSAWRMSLPEYKVSRLQACGDAMNLLGYR 359  
 DB 299 HLQTWVHNITRGKMGDAFHTNARDALNVSAWRMSLPEYKVSRLQACGDAMNLLGYR 359  
 QY 359 HVRSQEQQRNLDDLSTWTVEQI 385  
 DB 359 HVRSQEQQRNLDDLSTWTVEQI 385



PT Human and murine glycosyl sulfotransferase 3 and related  
 PT polynucleotides  
 PS Claim 2; Fig 1; 59pp; English.

CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of  
 CC the invention. The nucleic acid sequences, probes and primers derived  
 CC from these, proteins and antibodies are useful in detecting homologues.  
 CC The sequences, antibodies and methods are useful in the diagnosis and  
 CC treatment of diseases associated with selectin binding interactions,  
 CC including conditions associated with or resulting from the homing of  
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
 CC to secondary lymph organs.

XX Sequence 386 AA:

Query Match 100.0%; Score 2038; DB 20; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-205;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSWRSQSSFYVQ 60  
 DB 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSWRSQSSFYVQ 60  
 QY 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSYFDAYMEGPRRQ 120  
 DB 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSYFDAYMEGPRRQ 120  
 QY 121 SSLFQWENSALCSAPACDIIIPODEIIPRAHCRLLCSQPFVEYERAKRSYSHVYLKEVR 180  
 DB 121 SSLFQWENSALCSAPACDIIIPODEIIPRAHCRLLCSQPFVEYERAKRSYSHVYLKEVR 180  
 QY 181 FENLQSLYPLKDPSLNLIHVLVDRPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240  
 DB 181 FENLQSLYPLKDPSLNLIHVLVDRPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240  
 QY 241 QPYVMQVICOSELEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300  
 DB 241 QPYVMQVICOSELEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300  
 QY 301 HLOTWVHNITRKGMDGDAFHFNARDAALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360  
 DB 301 HLOTWVHNITRKGMDGDAFHFNARDAALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360  
 QY 361 HVSESEQRNLLDLSTWTVPEQIH 386  
 DB 361 HVSESEQRNLLDLSTWTVPEQIH 386

RESULT 2

AAM93309 ID AAM93309 standard; Protein; 386 AA.

XX AAM93309;

XX 06-NOV-2001 (first entry)

XX Human polypeptide, SEQ ID NO: 2817.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 93JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI: 2001-524255/58.

XX N-PSDB: AAK94229.

PS Claim 8; SEQ ID NO 2817; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a polypeptide  
 CC encoded by a full length human cDNA of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 386 AA:

Query Match 99.9%; Score 2035; DB 22; Length 386;  
 Best Local Similarity 99.7%; Pred. No. 3.1e-205;

Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSWRSQSSFYVQ 60  
 DB 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSWRSQSSFYVQ 60  
 QY 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSYFDAYMEGPRRQ 120  
 DB 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSYFDAYMEGPRRQ 120  
 QY 121 SSLFQWENSALCSAPACDIIIPODEIIPRAHCRLLCSQPFVEYERAKRSYSHVYLKEVR 180  
 DB 121 SSLFQWENSALCSAPACDIIIPODEIIPRAHCRLLCSQPFVEYERAKRSYSHVYLKEVR 180  
 QY 181 FENLQSLYPLKDPSLNLIHVLVDRPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240  
 DB 181 FENLQSLYPLKDPSLNLIHVLVDRPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240  
 QY 241 QPYVMQVICOSELEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300  
 DB 241 QPYVMQVICOSELEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300  
 QY 301 HLOTWVHNITRKGMDGDAFHFNARDAALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360  
 DB 301 HLOTWVHNITRKGMDGDAFHFNARDAALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360  
 QY 361 HVSESEQRNLLDLSTWTVPEQIH 386  
 DB 361 HVSESEQRNLLDLSTWTVPEQIH 386

RESULT 3

AAU11274 ID AAU11274 standard; Protein; 380 AA.

XX AAU11274;

XX 12-MAR-2002 (first entry)

XX Human L-selectin sulfotransferase-2 (LSST-2) protein.

XX Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;

XX L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 21:45:02 ; Search time 70 Seconds

(without alignments)  
734.782 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038  
Sequence: 1 MLLPKKKLLFLVVSOMATL.....EQRNLLDLSTWVPEQIH 386

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID2/gcgdata/geneseq/emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2038	100.0	386	20 AAY3918 Human glycosyl sul
2	2035	99.9	386	22 AAM93309 Human polypeptide,
3	2008	98.5	380	23 AAU11274 Human l-selectin s
4	1936	95.0	386	21 AAY79219 Human transferrase
5	1513.5	74.3	388	20 AAY3919 Mouse glycosyl sul
6	1203.5	59.2	418	21 AAB41947 Human ORFX ORF1711
7	1028	50.4	418	22 AAB81557 Mouse intestinal N
8	1019	50.0	395	22 AAY72640 Human glycosyl sul
9	1019	50.0	395	23 ABB81554 Human corneal N-ac
10	1019	50.0	395	23 AAB15438 Human drug metabol

11	1017.5	49.9	395	22 AAY72638
12	1017.5	49.9	395	23 AAU11275
13	1013.5	49.7	395	23 ABB81555
14	1008	49.5	390	22 AAY72639
15	1008	49.5	390	23 ABB81556
16	821	40.3	171	23 ABB81560
17	599.5	29.4	483	20 AAY31656
18	591.5	29.0	530	22 AAB93367
19	587.5	28.8	484	20 AAY31657
20	587.5	28.8	531	23 AA069414
21	549	26.9	411	19 AAM61100
22	519.5	25.5	169	23 ABB81558
23	517.5	25.1	169	23 ABB81559
24	500.5	24.6	458	18 AAM06480
25	482	23.7	479	19 AAM2863
26	336	16.5	169	23 ABB81561
27	335.5	16.5	179	23 ABB81562
28	293	14.4	174	23 ABB81563
29	210	10.3	363	22 ABB64512
30	186.5	9.2	183	22 ABB68582
31	109	5.3	1207	22 AAY72643
32	108.5	5.3	596	22 AAY72641
33	108.5	5.3	1222	22 AAY72642
34	100.5	4.9	775	23 ABB97942
35	96.5	4.7	791	23 AAB22917
36	94	4.6	3588	14 AAB34712
37	92.5	4.5	2354	22 ABB60511
38	92	4.5	359	20 AAY06628
39	92	4.5	359	21 AAY84309
40	91.5	4.5	5373	22 AAU14603
41	91.5	4.5	1847	21 AAY52002
42	91	4.5	1847	21 AAY51631
43	89.5	4.4	346	23 ABB81056
44	89	4.4	566	22 ABB3033
45	89	4.4	920	23 ABB25932

#### ALIGNMENTS

RESULT 1  
AAY3918  
ID AAY3918 standard; Protein: 386 AA.  
XX  
AC AAY3918;  
XX  
DT 08-DEC-1999 (first entry)  
XX  
DE Human glycosyl sulfotransferase-3 protein sequence.  
XX  
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
KW selectin binding interaction; inflammation; lymphocyte homing; human;  
KW secondary lymph organ.  
XX  
OS Homo sapiens.  
XX  
PN WO949018-A1.  
XX  
PD 30-SEP-1999.  
XX  
PF 26-FEB-1999; 99WO-US04316.  
XX  
PR 20-MAR-1998; 98US-0045284.  
PR 12-NOV-1998; 98US-0190911.  
PA (REGC ) UNIV CALIFORNIA.  
PA (SYNT ) SYNTAX USA INC.  
XX  
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;  
XX WPI, 1999-580442/49.  
XX DR N-PSDB; AA20792.  
XX



Db 105 TTGCTGCGGGCCGGCGCTTCACTCTGCTCCATCATGTT 64

RESULT 15  
US-09-815-242-7888

Sequence 7888, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

Prokaryotes

FILE REFERENCE: ELITRA.01A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 1410

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7888

LENGTH: 2838

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2838)

US-09-815-242-7888

Query Match 2.0%; Score 40; DB 10; Length 2838;

Best Local Similarity 44.8%; Pred. No. 0.31;

Matches 154; Conservative 0; Mismatches 190; Indels 0; Caps 0;

QY 634 CTGACAGCTCTCTGTCAGTCAAGCCCTTTGAGTGTGAGAGAGCCCTGCCCTCTCTA 693

Db 345 CGGTACCCCGCGCTGCGCGGCGACGACATCCGCTGAGCGCAGACCGTCAGCCAGAT 404

QY 694 CAGCCAGCTGCTGCTCAAGAGAGTGCCTTCTTCACTGCAATCCCTCTACCCGCTGCT 753

Db 405 GGTGACAGAGTCTCTGCGCTCGCGGAGAGGACAGCATGCTGCTGCGCGCGGTAT 464

QY 754 GAAGAGCCCTCTCTCAACCTGCAATGTCGACCTGTCCGGGAGCCCGGGCGCTGTT 813

Db 465 CCGGAGGCGCAAGGCGGACCTGCGCGGTGTCGACGAGATGCGCGCGAGGGCTTGT 524

QY 814 CCGTCCGAGAGACGACAAAGGAGATCTCATGATGATGACAGTGCATTTGATGGGCA 873

Db 525 CCGGCGCGGCTGCGCGGAGAGCTCTACGAACTGAGAGAGTGCAGAACTGATAGCA 584

QY 874 GCATGACGAGAACTCAAGAGAGAGACCAACCTACTATGTATGTCAGTCACTTCCA 933

Db 585 GAAGAGAGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 644

QY 934 AAGCAGCTGAGATCTACAGAGACCATCCATCTCTTGGCCAGG 977

Db 645 ACGCTGCGGAGATGCTGAGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 688

Search completed: January 10, 2003, 17:55:18  
Job time : 296 secs

**This Page Blank (uspto)**

QY	708	TCAAGAGAGTCCGCTTTCTTCAACTGCGAGTCCCTTACC	CGCTGCTGAAGACCCCTCC	767
Db	3	TAAAGAGTGTCCGGCTTTCGACGTGGGGGTCTGGCCCACTGCTGGAGACCGGGCC	62	
QY	768	TCAAGCTCATATCGTGCACCTGGTCCGGGACCCCGGGCCGCTGTTCCGTTCCGAGAAC	827	
Db	63	TGGACCTTAAGTATCTCCACTTGGTGGCTGATCCCCGGCGGTGGCGAGTTTCAGCGATCC	122	
QY	828	GCACAAAGGAGATCTCATGATTGACAGTCCGATTTGTATGGGGCAGCATGACGAC	883	
Db	123	GCTCCGCGCACGGCTCATCGTAGAGACCTACAGAGTGTGGCAGCGCAGACCG	178	
RESULT 14				
US-09-833-790-207/c				
; Sequence 207, Application US/09833790				
; Patent No. US20020068288A1				
GENERAL INFORMATION:				
APPLICANT: Lodes, Michael J.				
APPLICANT: Wang, Tongtong				
APPLICANT: Secrist, Heather				
APPLICANT: Monamach, Raodoh				
APPLICANT: Indrias, Carol Y.				
APPLICANT: Fan, Liqun				
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY				
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER				
FILE REFERENCE: 210121.512				
CURRENT APPLICATION NUMBER: US/09/833,790				
CURRENT FILING DATE: 2001-04-11				
NUMBER OF SEQ ID NOS: 440				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 207				
LENGTH: 395				
TYPE: DNA				
ORGANISM: Homo sapien				
US-09-833-790-207				
Query Match				
Best Local Similarity 2.1%; Score 42.8; DB 10; Length 395;				
Matches 110; Conservative 0; Mismatches 112; Indels 0; Gaps 0;				
QY	205	GCTCTGCTGTTTCTGTTTCCCAATGCGCATTTGCGCTATTTCTTCCACATGTACAG	264	
Db	285	GCCCAAGCCTTGTATGTCGAGATGTCATCTTGGGCACTTTGGC	226	
QY	265	CCACACATCAGTCCCTCTGATGAAGACACAGCCGAGCCGATGCTGCTTC	324	
Db	225	CCGCTCCCGCGGACACACATGAGGATTCATGCGCGCTTACCGCGGTCCGCGCT	166	
QY	325	GTCCTCTGACCGCTCTGCTCTTTTGTGGGCGACGTTTTTGGGACACCCCAATGT	384	
Db	165	GTTTTCTGTTTCCCGCGGCCCGCGCGGTGGAGATTGCCGCCCGCCCGCAAGT	106	
QY	385	TTTCTACGTGATGACCGCGCTGGACGTTGTGATGACCTT	426	

Db 1711 TCACCTTCACGAGATCAACAGAGGTGGAGAGATTGCTACAGCCATGCGCCGCTGG 1770  
Qy 1254 GCTACCGGACGTCA 1268  
Db 1771 GCTATGACCGGCTCA 1785

## RESULT 9

US-09-735-705-63  
Sequence 63, Application US/09735705  
Patent No. US20020052329A1  
GENERAL INFORMATION:

APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hosken, Nancy  
APPLICANT: Fanger, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Henderson, Robert A.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.455C14  
CURRENT APPLICATION NUMBER: US/09/735.705  
CURRENT FILING DATE: 2000-12-12  
NUMBER OF SEQ ID NOS: 419  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 63  
LENGTH: 731  
TYPE: DNA  
ORGANISM: Homo sapien

NAME/KEY: misc.feature  
LOCATION: (1) (731)  
OTHER INFORMATION: n = A,T,C or G  
US-09-735-705-63

Query Match 3.1%; Score 62.6; DB 10; Length 731;  
Best Local Similarity 58.2%; Pred. No. 2.8e-08;  
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 702 TGGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCGCTCTTCAACCGCTGCTGAAGACC 761  
Db 3 TAGTCATAAAGGGTGTGCGGTCTTTCACGCTGGCGGCTTTGGCGCCACTGCTCGAGAGCC 62  
Qy 762 CCTCCCTCAACCTCATATCGTGTGACCTGGTCCGGAGACCCCGCGCGCTTCCGTTCCC 821  
Db 63 CGGCGCTGGAGCTCAAGCTCATCTGCTGTGCTGTATCCCGCGGCTGCGAGTTCCAC 122  
Qy 822 GAGAAGCACAAGAGGAGATCTCATGTATGACAGTCCATTTGATGGGGCAGCAGATGAGC 881  
Db 123 GGATCCCGCTGGCGCAGCGCTCATCGGTGAGAGCTACAGGTGGTGGCAGCGGAGACC 182  
Qy 882 AGAACTCA 890  
Db 183 GCGAGCTCA 191

RESULT 10  
US-09-850-716A-63  
Sequence 63, Application US/09850716A  
Patent No. US20020115139A1  
GENERAL INFORMATION:

APPLICANT: Kalos, Michael D.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Retter, Marc W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.455C15  
CURRENT APPLICATION NUMBER: US/09/850.716A  
CURRENT FILING DATE: 2001-05-07  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 63  
LENGTH: 731  
TYPE: DNA  
ORGANISM: Homo sapien

NAME/KEY: misc.feature  
LOCATION: (1) (731)  
OTHER INFORMATION: n = A,T,C or G  
US-09-850-716A-63

Query Match 3.1%; Score 62.6; DB 10; Length 731;  
Best Local Similarity 58.2%; Pred. No. 2.8e-08;  
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 702 TGGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCGCTCTTCAACCGCTGCTGAAGACC 761  
Db 3 TAGTCATAAAGGGTGTGCGGTCTTTCACGCTGGCGGCTTTGGCGCCACTGCTCGAGAGCC 62  
Qy 762 CCTCCCTCAACCTCATATCGTGTGACCTGGTCCGGAGACCCCGCGCGCTTCCGTTCCC 821  
Db 63 CGGCGCTGGAGCTCAAGCTCATCTGCTGTGCTGTATCCCGCGGCTGCGAGTTCCAC 122  
Qy 822 GAGAAGCACAAGAGGAGATCTCATGTATGACAGTCCATTTGATGGGGCAGCAGATGAGC 881  
Db 123 GGATCCCGCTGGCGCAGCGCTCATCGGTGAGAGCTACAGGTGGTGGCAGCGGAGACC 182  
Qy 882 AGAACTCA 890  
Db 183 GCGAGCTCA 191

## RESULT 11

US-09-897-778-63  
Sequence 63, Application US/09897778  
Patent No. US20020147143A1  
GENERAL INFORMATION:

APPLICANT: Wang, Tongtong  
APPLICANT: Margerakis, Margarita  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darlick  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Henderson, Robert A.  
APPLICANT: Peckham, David W.  
APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.455C16  
CURRENT APPLICATION NUMBER: US/09/897.778  
CURRENT FILING DATE: 2001-06-28  
NUMBER OF SEQ ID NOS: 467  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 63  
LENGTH: 731  
TYPE: DNA  
ORGANISM: Homo sapiens

NAME/KEY: misc.feature  
LOCATION: 237, 248, 263, 288, 312, 317, 323, 326, 337, 352, 362, 370,  
LOCATION: 377, 400, 411, 414, 434, 436, 445, 457, 473, 486, 497, 498,  
LOCATION: 502, 512, 531, 546, 554, 563, 566, 588, 597, 608, 611,  
LOCATION: 613, 615, 627, 632, 640, 641, 644, 654, 660, 663, 665  
OTHER INFORMATION: n = A,T,C or G  
NAME/KEY: misc.feature  
LOCATION: 671, 678, 692, 697, 698, 699, 704, 705, 712, 714, 717, 718,  
LOCATION: 719, 722, 725, 730, 731  
OTHER INFORMATION: n = A,T,C or G  
US-09-897-778-63



RESULT 6  
US-09-927-602-38  
; Sequence 38, Application US/09927602

Query Match	16.6%	Score 337.8	DB 10	Length 48436
Best Local Similarity	65.1%	Pred. No. 4e-88		
Matches 514	Conservative	0	Mismatches 272	Indels 3
			Gaps	1
QY 302	GAGCGATGACAGTGTGCTTGTCTCTCTGCGGCGCTTGACCTCTCTTTTGTGGGACAG	361		
Db 47651	GCGCGGTGCATGTGTGTGTGTCTCTCTGCGGCGCTTGACCTCTCTTTGTGGGACAG	47710		
QY 362	CTTTTGGGAGACACCATGTTTCTTACCGTGAAGAGAGCCCGCTGCGAGCTGTGATG	421		
Db 47711	CTCTTACACGAGACCCCGAGCTCTTCTTACCTTAAGAGAGCCCGGTGTGCGACGTGTGACC	47770		
QY 422	ACCTTGAAGCAGACACCGCCTGTGATGTGCATATGCTGTGCGATCTGATACCGGCC	481		
Db 47771	ACCTGTGCGAGGCGAGCGCCGCAACGCTGCATATGCTGTGCGCGACCTGTGTGCCCTCC	47830		
QY 482	GTCCTCTTGGGACATGAGCTCTTTTATGCTTCATGAACTGTGTCCCGGAGACAG	541		
Db 47831	GTCCTCTCTGGGACATGAGCTCTTTTATGCTTCATG---CTTGGCGCCCAACCTG	47887		
QY 542	TTCAGCCTCTTTCAGTGGGAGAACAGCGCGGCCCTGTGTCTGTGCACCTGTGCTGATC	601		
Db 47888	TCCGACCTCTTCAGTGGGCGCCCTGAGCCGTGACCTGTGTCTGTGCACCGCTGTAGTGC	47947		
QY 602	ATCCCAAGATGAATCATCTCCCGGCTCATGTCCAGGCTCCTGTGCACTACAGGCC	661		
Db 47948	TTTCCCGAGGCGCATCAGCAGCAGGCGCTGTGTGCAAGCACATGTGCGCGCGAGTCC	48007		
QY 662	TTTGAGTGTGTGAGAGGCGCTGCGGCTCTCAGACGCAGCTGGGCTCAAGAGGTGGCC	721		
Db 48008	TTTACCTTGCGCGGAGGCGCTGCGGCTCTCTCAGACGCAGCTGTGTGCTCAAGAGGTGGCC	48067		
QY 722	TTCTTCAACCTGAGTCCCTTACCCGCTGTGAAAGCCCTCTCCCTCACTGCATATC	781		
Db 48068	TTCTTCAACCTGAGTCCCTTACCCGCTGTGAAAGCCCTCTCCCTCACTGCATATC	48127		
QY 782	GTCACCTGTGTCGGGACCCCGGCGCTGTCTCGTTCCCGAAGAGCCCAAGGAGAT	841		
Db 48128	GTCACCTGTGTCGGGACCCCGGCGCTGTCTCGCTCCCGGAGAGACAGCCAGAGT	48187		
QY 842	CTCATGATGACAGTGTGATGTGATGGGAGCATGACAGAACTCAAGAGAGAGAC	901		
Db 48188	CTGAGCGCTGACAGAGGATGTGTGTGGGACCAAGGCGACGTGGATGGAGGCCACCC	48247		
QY 902	CACCCCTACTATGTATGTACAGTATCTGCCAAGCCAGCTGAGATCTTACAGACCATC	961		
Db 48248	GGCCTGTGCGGTGTGCGGAGGTGTGCCGTAAGCCAGCTAGCATTCGCCAGGCGCCACA	48307		
QY 962	CAGTCTTTGCCAAGGCGCTCAGAGAGCGTACTGTGTGGGTATGAGACCGTGGT	1021		
Db 48308	CTCAAGCGGACACCTTTCTGTGGGCGGCTACCGCTGTGTGTGCTTCAAGAGACTGTGG	48367		
QY 1022	CGAGCCCTGTGGCCAGACTTCCCGAATGTATGATTTGTGGGATTTGAATCTTGGCC	1081		

```

1  RESULT 4
2  US-09-998-2595
3  : Sequence 2595, Application US/09998598
4  : Patent No. US20020150922A1
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Stolk, John A.
9  : APPLICANT: Xu, Jiaqun
10 : APPLICANT: Chenault, Ruth A.
11 : APPLICANT: Meagher, Madelein Joy
12 :
13 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
14 : FILE REFERENCE: 210121.561
15 :
16 : CURRENT APPLICATION NUMBER: US/09/998,598
17 :
18 : CURRENT FILING DATE: 2001-11-16
19 :
20 : NUMBER OF SEQ ID NOS: 2606
21 :
22 : SOFTWARE: Corlixa Invention Disclosure Database
23 :
24 : SEQ ID NO 2595
25 :
26 : LENGTH: 505

```

Best Local Similarity 64.  
Matches 654: Conservative

Query Match	20.48;	Score 414.4;	DB 10;	Length 2544;
Best Local Similarity	64.5%;	Pred. No. 1.6e-11;		
Matches 654;	Conservative	0;	Mismatches 351;	Indels 9;
				Gaps 2

Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTGGAGGCGAGATGGCTTCAGTCTGGGAAATGCTTCATTTGGTTCCTCCAG 60  
 Db 1 GCGTGGAGGCGAGATGGCTTCAGTCTGGGAAATGCTTCATTTGGTTCCTCCAG 60  
 QY 61 CCCACCTCAAGAGCTCTCCCAACCCCTTGAATCAGCAGTAAAGCTTTACTTTCA 120  
 Db 61 CCCACCTCAAGAGCTCTCCCAACCCCTTGAATCAGCAGTAAAGCTTTACTTTCA 120  
 QY 121 CAGCTTCCTGGGAGGAGTGGTTCCTCAAGCCCTTGGCAAGTCTTCCACTCAGAC 180  
 Db 121 CAGCTTCCTGGGAGGAGTGGTTCCTCAAGCCCTTGGCAAGTCTTCCACTCAGAC 180  
 QY 181 AATGCTACTGGCTTAAATAAATGAAGCTCCTGCTTTCTGATTCAGATGAGCATCTT 240  
 Db 181 AATGCTACTGGCTTAAATAAATGAAGCTCCTGCTTTCTGATTCAGATGAGCATCTT 240  
 QY 241 GGGCTATATCTTCACAGTACAGCAGCAGACATAGCTCCTGCTTCTATGAAGCAGCC 300  
 Db 241 GGGCTATATCTTCACAGTACAGCAGCAGACATAGCTCCTGCTTCTATGAAGCAGCC 300  
 QY 301 CGAGGCGATGCGAGCTGCTGCTCTCTCTGCGGCTCTGCTCTTCTTTGGGGCA 360  
 Db 301 CGAGGCGATGCGAGCTGCTGCTCTCTCTGCGGCTCTGCTCTTCTTTGGGGCA 360  
 QY 361 GCTTTTGGGAGCAGCAGCAGATGTTTCTTACCTGATGAGAGCCCTGAGATGAGAT 420  
 Db 361 GCTTTTGGGAGCAGCAGCAGATGTTTCTTACCTGATGAGAGCCCTGAGATGAGAT 420  
 QY 421 GACCTTCAGAGAGCAGCAGCAGCTGATGCTGACATGAGTGGGAGATGATGAGGAG 480  
 Db 421 GACCTTCAGAGAGCAGCAGCAGCTGATGCTGACATGAGTGGGAGATGATGAGGAG 480  
 QY 481 GCTCTCTCTTGGCAGCATGAGCGCTTGTGATGCTCAGTGAAGCTGCTCCCGGAGCA 540  
 Db 481 GCTCTCTCTTGGCAGCATGAGCGCTTGTGATGCTCAGTGAAGCTGCTCCCGGAGCA 540  
 QY 541 GTCCAGCCTCTTTCAGTGGGAGAGCAGCGGCGCTGCTTTCAGCAGCTCCCTGAGAT 600  
 Db 541 GTCCAGCCTCTTTCAGTGGGAGAGCAGCGGCGCTGCTTTCAGCAGCTCCCTGAGAT 600  
 QY 601 CATCCCAAGAGATGAATCATCCCGCGGCTCACTGAGGCTCCTGCGAGTCAACAGCC 660  
 Db 601 CATCCCAAGAGATGAATCATCCCGCGGCTCACTGAGGCTCCTGCGAGTCAACAGCC 660  
 QY 661 CTTTGGAGTGGGAGAGGCTGCGGCTGCTTACAGCAGCAGTGGTCTCAAGAGAGTGGC 720  
 Db 661 CTTTGGAGTGGGAGAGGCTGCGGCTGCTTACAGCAGCAGTGGTCTCAAGAGAGTGGC 720  
 QY 721 CTTCTTCAACTGAGTCCCTTACCGGCTGCTGAAGAGACCCCTCCCTCAACTGATAT 780  
 Db 721 CTTCTTCAACTGAGTCCCTTACCGGCTGCTGAAGAGACCCCTCCCTCAACTGATAT 780  
 QY 781 CGTGCACCTGCTCGGAGACCCCGGCGCTGCTCCGAGAGACGCAAGAGGAGAG 840  
 Db 781 CGTGCACCTGCTCGGAGACCCCGGCGCTGCTCCGAGAGACGCAAGAGGAGAG 840  
 QY 841 TCTCATGATGAGAGTGGATGATGATGGGAGCAGATGAGCAGAACTCAAGAGAGAG 900  
 Db 841 TCTCATGATGAGAGTGGATGATGATGGGAGCAGATGAGCAGAACTCAAGAGAGAG 900  
 QY 901 CCAACCTTACTATGATGAGTGCATCTGCGCAAGCCAGCTGAGATCTAGAGACAT 960  
 Db 901 CCAACCTTACTATGATGAGTGCATCTGCGCAAGCCAGCTGAGATCTAGAGACAT 960  
 QY 961 CCAAGTCTTCCCAAGGCGCTGAGAGAGCTGCTGCTTGGCGTATGAGAGCTGGC 1020  
 Db 961 CCAAGTCTTCCCAAGGCGCTGAGAGAGCTGCTGCTTGGCGTATGAGAGCTGGC 1020  
 QY 1021 TCGAGGCGCTGCTGGCGCAGACTTCCCGAATGATGAATTCGTGGATTCCTGGC 1080  
 Db 1021 TCGAGGCGCTGCTGGCGCAGACTTCCCGAATGATGAATTCGTGGATTCCTGGC 1080

QY 1081 CCATCTTCAGACCTGGGTGCTATACATCACCCAGGCGATGGGTGACACGCTTT 1140  
 Db 1081 CCATCTTCAGACCTGGGTGCTATACATCACCCAGGCGATGGGTGACACGCTTT 1140  
 QY 1141 CCACACAAATGCCAGGAGTCCCTTAATGCTCCAGAGCTTGGCGCTTGGCTTGA 1200  
 Db 1141 CCACACAAATGCCAGGAGTCCCTTAATGCTCCAGAGCTTGGCGCTTGGCTTGA 1200  
 QY 1201 TGAAGAGTCTTCAGACTTCAAGAGCTGCTGAGATGAGATGAGATGAGATGAG 1260  
 Db 1201 TGAAGAGTCTTCAGACTTCAAGAGCTGCTGAGATGAGATGAGATGAGATGAG 1260  
 QY 1261 CCAGCTGAGTTCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 1261 CCAGCTGAGTTCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 QY 1321 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAGAGCTTGGTCCACCTGCTGAGCC 1380  
 Db 1321 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAGAGCTTGGTCCACCTGCTGAGCC 1380  
 QY 1381 TCACTGACTTCTCTGAAGCTTCTGAGGCTTGGCTTACATCTCTGAGCTTAACTATG 1440  
 Db 1381 TCACTGACTTCTCTGAAGCTTCTGAGGCTTGGCTTACATCTCTGAGCTTAACTATG 1440  
 QY 1441 TCTGTGGATTCACACTGAGTGAATGCTGCTCAGCAGCTGCTCAAGCAGAGACTTT 1500  
 Db 1441 TCTGTGGATTCACACTGAGTGAATGCTGCTCAGCAGCTGCTCAAGCAGAGACTTT 1500  
 QY 1501 GTGTCCATGCTTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 Db 1501 GTGTCCATGCTTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 QY 1561 CAGTGAAGAGGATGATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620  
 Db 1561 CAGTGAAGAGGATGATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620  
 QY 1621 AGACTTGTGGCTGAGAGGCTTATTAAGCAGCAGCAGATGAGATGAGATGAGAT 1680  
 Db 1621 AGACTTGTGGCTGAGAGGCTTATTAAGCAGCAGCAGATGAGATGAGATGAGAT 1680  
 QY 1681 ACCCTCCCTGCAACTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 Db 1681 ACCCTCCCTGCAACTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 QY 1741 ATTTTCCAGAGATGCAAAATCTGAGGCTTGAAGTCCAGTGGATTCAGAGAGAG 1800  
 Db 1741 ATTTTCCAGAGATGCAAAATCTGAGGCTTGAAGTCCAGTGGATTCAGAGAGAG 1800  
 QY 1801 GTGGAGAGAGGTTGAGATGCTTATGAGCTTACCATCAGACTATGCTATCAGA 1860  
 Db 1801 GTGGAGAGAGGTTGAGATGCTTATGAGCTTACCATCAGACTATGCTATCAGA 1860  
 QY 1861 AATATGAACAAATCTGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
 Db 1861 AATATGAACAAATCTGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
 QY 1921 TGCATTTGAATATCACTTCCCTGAGCTTTCATGAGTTCATGAGAGAGCTTGA 1980  
 Db 1921 TGCATTTGAATATCACTTCCCTGAGCTTTCATGAGTTCATGAGAGAGCTTGA 1980  
 QY 1981 AAGCTGCACTGTTAATACTAAATTCCTCAATAGAGAGAGAGAGAGAGAGAG 2032  
 Db 1981 AAGCTGCACTGTTAATACTAAATTCCTCAATAGAGAGAGAGAGAGAGAGAG 2032

RESULT 3  
 US-09-919-580-194  
 : Sequence 194, Application us/09919580  
 : Patent No. US20020110832A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Pyile, Ruth  
 : APPLICANT: Xu, Jiangchun



QY	1321	TGTCCTTGAGCAATACACATACAGGCTTGAGAGGCTTGTCTGCACCTGGTGTACAC	1380	
Db	1321	TGTCCTTGAGCAATACACATACAGGCTTGAGAGGCTTGTCTGCACCTGGTGTACAC	1380	
QY	1381	TCAGTCACTTTCCTGTGAATGCTTCTAGACCTTTCCTACATCTCTGAGCCTTA	1440	
Db	1381	TCAGTCACTTTCCTGTGAATGCTTCTAGACCTTTCCTACATCTCTGAGCCTTA	1440	
QY	1441	TCGTGGGTATCCACACTGAGTGTGAGTGTGTCCACAGTGGCTCAAGCAGAGACTTT	1500	
Db	1441	TCGTGGGTATCCACACTGAGTGTGAGTGTGTCCACAGTGGCTCAAGCAGAGACTTT	1500	
QY	1501	GTGTCCATGCTTGTGTCTAGAAAAACAGACTGGGGAACTTATGTGACGACATCCAC	1560	
Db	1501	GTGTCCATGCTTGTGTCTAGAAAAACAGACTGGGGAACTTATGTGACGACATCCAC	1560	
QY	1561	CAGTGAACACAGGATATGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1620	
Db	1561	CAGTGAACACAGGATATGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1620	
QY	1621	AGACTTTGTGGCTGAGAGGCTATTAAAGCAGACAGATCAGTGTGATGATTCATTA	1680	
Db	1621	AGACTTTGTGGCTGAGAGGCTATTAAAGCAGACAGATCAGTGTGATGATTCATTA	1680	
QY	1681	ACCTCCCTGTCCATCTTCCCAATGGGAAATGATCTTTTCCAAAGAGCTCACACAC	1740	
Db	1681	ACCTCCCTGTCCATCTTCCCAATGGGAAATGATCTTTTCCAAAGAGCTCACACAC	1740	
QY	1741	ATTTCCTCAGAGATGCAAAATTCGAGCCCTTGAGTTCCTCAGTGTATTCAGAGAA	1800	
Db	1741	ATTTCCTCAGAGATGCAAAATTCGAGCCCTTGAGTTCCTCAGTGTATTCAGAGAA	1800	
QY	1801	GTGGGACACAGTGTGATTCCTTACTTATAGCTTGTAGCATCACAAGTATCCGTAATCA	1860	
Db	1801	GTGGGACACAGTGTGATTCCTTACTTATAGCTTGTAGCATCACAAGTATCCGTAATCA	1860	
QY	1861	AATATGAACAAATCTCTGACAAAAAGACAAGCTCTTAAGTTCCAGAGTCCCTGGGC	1920	
Db	1861	AATATGAACAAATCTCTGACAAAAAGACAAGCTCTTAAGTTCCAGAGTCCCTGGGC	1920	
QY	1921	TGCATTTAATATACCTTCCCTCTGATTTTCCATCAATAGAGACCTTGAAGCTGTG	1980	
Db	1921	TGCATTTAATATACCTTCCCTCTGATTTTCCATCAATAGAGACCTTGAAGCTGTG	1980	
QY	1981	AAGCTGCATCTGTTAATACTAAATTCACCAATAGAGAAAAA	2032	
Db	1981	AAGCTGCATCTGTTAATACTAAATTCACCAATAGAGAAAAA	2032	
RESULT 2				
US-09-816-825-1				
Sequence 1, Application US/09816825				
Patent No. US20010051370A1				
GENERAL INFORMATION:				
APPLICANT: Bistrup, Annette				
APPLICANT: Rosen, Steven D				
APPLICANT: Hemmerich, Stefan				
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3				
FILE REFERENCE: 6510-107CON				
CURRENT APPLICATION NUMBER: US/09/816,825				
CURRENT FILING DATE: 2001-03-22				
PRIOR APPLICATION NUMBER: 09/045,284				
PRIOR FILING DATE: 1998-03-20				
NUMBER OF SEQ ID NOS: 9				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 1				
LENGTH: 2032				
TYPE: DNA				
ORGANISM: Homo sapiens				
US-09-816-825-1				
Query Match	100.0%	Score 2032:	DB 10:	Length 2032:

Db 181 AATGCTACTGCTAAATAAATGAAGCTCTGCTGTTCTGTTTCCAGATGCCATCTT 240

1	EARLIER APPLICATION NUMBER: 60/0443,671
2	EARLIER FILING DATE: 1997-04-11
3	EARLIER APPLICATION NUMBER: 60/0443,674
4	EARLIER FILING DATE: 1997-04-11
5	EARLIER APPLICATION NUMBER: 60/043,669
6	EARLIER FILING DATE: 1997-04-11
7	EARLIER APPLICATION NUMBER: 60/043,312
8	EARLIER FILING DATE: 1997-04-11
9	EARLIER APPLICATION NUMBER: 60/043,313
10	EARLIER FILING DATE: 1997-04-11
11	EARLIER APPLICATION NUMBER: 60/043,315
12	EARLIER FILING DATE: 1997-04-11
13	EARLIER APPLICATION NUMBER: 60/043,315
14	EARLIER FILING DATE: 1997-04-11
15	EARLIER APPLICATION NUMBER: 60/048,974
16	EARLIER FILING DATE: 1997-06-06
17	EARLIER APPLICATION NUMBER: 60/056,886
18	EARLIER FILING DATE: 1997-08-22
19	EARLIER APPLICATION NUMBER: 60/056,887
20	EARLIER FILING DATE: 1997-08-22
21	EARLIER APPLICATION NUMBER: 60/056,889
22	EARLIER FILING DATE: 1997-08-22
23	EARLIER APPLICATION NUMBER: 60/056,893
24	EARLIER FILING DATE: 1997-08-22
25	EARLIER APPLICATION NUMBER: 60/056,530
26	EARLIER FILING DATE: 1997-08-22
27	EARLIER APPLICATION NUMBER: 60/056,878
28	EARLIER FILING DATE: 1997-08-22
29	EARLIER APPLICATION NUMBER: 60/056,662
30	EARLIER FILING DATE: 1997-08-22
31	EARLIER APPLICATION NUMBER: 60/056,872
32	EARLIER FILING DATE: 1997-08-22
33	EARLIER APPLICATION NUMBER: 60/056,882
34	EARLIER FILING DATE: 1997-08-22
35	EARLIER APPLICATION NUMBER: 60/056,637
36	EARLIER FILING DATE: 1997-08-22
37	EARLIER APPLICATION NUMBER: 60/056,903
38	EARLIER FILING DATE: 1997-08-22
39	EARLIER APPLICATION NUMBER: 60/056,888
40	EARLIER FILING DATE: 1997-08-22
41	EARLIER APPLICATION NUMBER: 60/056,879
42	EARLIER FILING DATE: 1997-08-22
43	EARLIER APPLICATION NUMBER: 60/056,880
44	EARLIER FILING DATE: 1997-08-22
45	EARLIER APPLICATION NUMBER: 60/056,894
46	EARLIER FILING DATE: 1997-08-22
47	EARLIER APPLICATION NUMBER: 60/056,911
48	EARLIER FILING DATE: 1997-08-22
49	EARLIER APPLICATION NUMBER: 60/056,636
50	EARLIER FILING DATE: 1997-08-22
51	EARLIER APPLICATION NUMBER: 60/056,874
52	EARLIER FILING DATE: 1997-08-22
53	EARLIER APPLICATION NUMBER: 60/056,910
54	EARLIER FILING DATE: 1997-08-22
55	EARLIER APPLICATION NUMBER: 60/056,864
56	EARLIER FILING DATE: 1997-08-22
57	EARLIER APPLICATION NUMBER: 60/056,631
58	EARLIER FILING DATE: 1997-08-22
59	EARLIER APPLICATION NUMBER: 60/056,845
60	EARLIER FILING DATE: 1997-08-22
61	EARLIER APPLICATION NUMBER: 60/056,892
62	EARLIER FILING DATE: 1997-08-22
63	EARLIER APPLICATION NUMBER: 60/057,761
64	EARLIER FILING DATE: 1997-08-22
65	EARLIER APPLICATION NUMBER: 60/047,595
66	EARLIER FILING DATE: 1997-05-23
67	EARLIER APPLICATION NUMBER: 60/047,599
68	EARLIER FILING DATE: 1997-05-23
69	EARLIER APPLICATION NUMBER: 60/047,588
70	EARLIER FILING DATE: 1997-05-23
71	EARLIER APPLICATION NUMBER: 60/047,585
72	EARLIER FILING DATE: 1997-05-23
73	EARLIER APPLICATION NUMBER: 60/047,586

EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11	EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11	EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23	EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11	EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06	EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05	EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22	EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05	EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13	EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02	

Search completed: January 10, 2003, 17:50:51  
Job time : 124 secs

; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,586

**This Page Blank (uspto)**

QY 370 GCAGCACCACGATGTTTCTACCT 393  
Db 1394 YYYYYYYYYYYYYYYYYYYY 1417

RESULT 14  
US-09-199-637A-168  
Sequence 168, Application US/09199637A  
Patent No. 6355411  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick  
APPLICANT: Goodman, Howard M.  
APPLICANT: Rahme, Laurence G.  
APPLICANT: Mahajan-Miklos, Shalina  
APPLICANT: Tan, Man-Wah  
APPLICANT: Drenkard, Eliana  
APPLICANT: Tsongalis, John  
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
FILE REFERENCE: 00786/361002  
CURRENT APPLICATION NUMBER: US/09/199,637A  
CURRENT FILING DATE: 1998-11-25  
PRIOR APPLICATION NUMBER: 60/066,517  
PRIOR FILING DATE: 1997-11-25  
NUMBER OF SEQ. ID NOS: 437  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 168  
LENGTH: 3010  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-168

Query Match 1.98; Score 37.6; DB 4; Length 3010;  
Best Local Similarity 44.08; Pred. No. 0.8; Mismatches 0;  
Matches 160; Conservative 0; Indels 0; Gaps 0;

QY 644 CTGTGACGTCAACAGCCCTTTGAGTGGAGAGGCTGCCGCTCTTACAGCCACGTG 703  
Db 1948 CTGTGGGCGCCAGTACAGAGCGTGTGCGGAGAGGCGCGAGACCAAGGCGCTGCGG 2007

QY 704 GTGCTCAAGAGGTGCGCTTTTAACTGCACTCCCTCTACCCGCTGCTGAANAACCC 763  
Db 2008 GCCCTGGCGGCTGCGCTGAGACCCCTGACCGCCGATGCCGCTGTGGTAC 2067

QY 764 TCCCTCAACCTGATGTCATGTCGATGTCGAGCCCGCGGCGTGTCCGTTCCGA 823  
Db 2068 ACCGCGCTGCTGCGCGATGTCGCGCGCCGCGAGAGCGTGGCGCGAGGGCGTGGCTG 2127

QY 824 GAAGCACAAGGAGATCTCATATTGACAGTGCATTTGTGATGGGCGACATGAGCAG 883  
Db 2128 TACCGCACCGAAGTGCCTGATGATACACCGCTTCCCAACGAGAGAAACAGCTG 2187

QY 884 AAACGCAAGAGGAGGAGCAACCTACTATGTGATGCAAGGCGATCTCCAAAGCCACCTG 943  
Db 2188 GCGATCTACCCGCGAGCGAGCTAGTGCCTTCCACCCGCTGCCGCTGACATGCGACCCCTG 2247

QY 944 GAGATCTACAAGACCATCCATCTTGTCCCAAGGCGCTGAGAGAACCTCACTGCTTGTG 1003  
Db 2248 GATATCGGCGCGACAAAGGCGCTGTCTACTTCCGATCAAGAGAAACACCGCTTCTC 2307

QY 1004 CGCT 1007  
Db 2308 GGCT 2311

RESULT 15  
US-09-149-476-296  
Sequence 296, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins

FILE REFERENCE: P2002P1  
CURRENT APPLICATION NUMBER: US/09/149,476  
EARLIER FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,584  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,500  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,587  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,492  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,598  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,613  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,582  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,596  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11

Db 3 TAGTCATTAAGGCGTGTGCGGCTTTTCAGACGTGGCGGCTTTGGCCCACTGCTGCGAGACC 62  
QY 762 CCTCCCTCAACCTGATATCGTGCACCGNGTCGGGACCCCGGCGCGTTCGGTTCC 821  
Db 63 CGGCCCTGAGACCTCAAGATCATCTGAGTGGTGTATCCCGCGCGGTGCGAGTTTCAC 122  
QY 822 GAGAACCGACAAAGGAGATCTCATGTATGTACACTCCCATTTGATGAGGCGAGATAGC 881  
Db 123 GGATCCGCTCGCGCCACGCGCTCATCCGTAGAGCCTACAGGTGGTGTGCGAGCGAGACC 182  
QY 882 AGAACTCA 890  
Db 183 GCGAGCTCA 191

RESULT 12  
US-09-643-597-63  
; Sequence 63, Application US/09643597  
; Patent No. 6426072  
; GENERAL INFORMATION:  
; APPLICANT: Mang, Tongtong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Bangun, Chaltanya S.  
; APPLICANT: Hosken, Nancy  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Mang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Menell, Patricia D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.455C11  
; CURRENT APPLICATION NUMBER: US/09/643,597  
; CURRENT FILING DATE: 2000-08-21  
; NUMBER OF SEQ ID NOS: 369  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 63  
; LENGTH: 731  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(731)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-643-597-63

Query Match 3.1%; Score 62.6; DB 4; Length 731;  
Best Local Similarity 58.2%; Pred. No. 1.9e-08;  
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGGTGTCAAGGAGGAGTGCCTTCTTCAACCTGAGTCCCTCTACCGCGTGTGAAAGAC 761  
Db 3 TAGTCATTAAGGCGTGTGCGGCTTTTCAGACGTGGCGGCTTTGGCCCACTGCTGCGAGACC 62  
QY 762 CCTCCCTCAACCTGATATCGTGCACCGNGTCGGGACCCCGGCGCGTTCGGTTCC 821  
Db 63 CGGCCCTGAGACCTCAAGATCATCTGAGTGGTGTATCCCGCGCGGTGCGAGTTTCAC 122  
QY 822 GAGAACCGACAAAGGAGATCTCATGTATGTACACTCCCATTTGATGAGGCGAGATAGC 881  
Db 123 GGATCCGCTCGCGCCACGCGCTCATCCGTAGAGCCTACAGGTGGTGTGCGAGCGAGACC 182  
QY 882 AGAACTCA 890  
Db 183 GCGAGCTCA 191

RESULT 13  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367

; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELERX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PT29pt-F15  
US-08-232-463-14

Query Match 2.8%; Score 56.4; DB 1; Length 7218;  
Best Local Similarity 4.7%; Pred. No. 4.3e-06;  
Matches 18; Conservative 215; Mismatches 151; Indels 0; Gaps 0;

QY 10 CCAGATGCGCTCCAGTCTGGCGGAAATGCTTCTCATTTGCTTCCAGCCACCTCA 69  
Db 1034 CGAGCTGTGCTCCAGGTGAGGTGAGGAGCTTGCATTTTTTTTTTTTTTTTTT 1093  
QY 70 ACAGCTCTCCCAACCCCTTGAGTCTCAGCAGTGTAAAGCTGTACTTTCACAGTTCT 129  
Db 1094 TT 1153  
QY 130 GAGAGGAGTGTCTTCTCAAGCCGCTTGCAGGCTTCCACTCAGACAAATGCTACT 189  
Db 1154 TT 1213  
QY 190 GCTTAAAAAAGAACTCTGCTGTTTGTGTTCCAGATGCCAATGCTGCTT 249  
Db 1214 TT 1273  
QY 250 CTTCACATGTACAGCACAAATCAGTCCGCTGTGTAAGGACAGACCGAGCGAT 309  
Db 1274 TT 1333  
QY 310 GCAGCTGCTGTCTGTCTTCTGCGCGCTGCTCTTCTTTGTGGGACAGTTTGG 369  
Db 1334 TT 1393

```
LOCATION: (411)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (414)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (434)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (436)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (445)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (457)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (473)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (486)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (497)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (498)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (502)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (512)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (531)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (546)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (554)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (563)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (565)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (588)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (597)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (606)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (611)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (613)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (615)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (627)
```

```
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (632)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (640)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (641)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (644)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (654)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (660)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (663)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (665)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (671)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (678)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (692)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (697)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (698)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (699)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (704)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (705)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (712)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (714)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (717)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (718)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (719)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (723)
```

Query Match 3.1%; Score 62.6; DB 4; Length 731;  
Best Local Similarity 58.2%; Pred. No. 1.9e-08;  
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGGTGCACAGAGGTGGCGTTCTTCACCTGACGTCCCTACCGCGTGGTGAAGACC 761

Db 1174 GTGCCACCTCAGCCCTGGGACGCTTGGAGAGAGAGTGGCAGATCCGACAGGTGT 1233  
QY 706 GCTCAGAGAGTGGCTCTTTCACCTGACCTCCCTTACCCGCTGCTGAAGACCCCTC 765  
Db 1234 TATCAAGGGCGGTGGCTTTCGATGTGTGTGGCGGCTGCTTAAGATCCAGC 1293  
QY 766 CCTCAGCTGATATGTCAGCTGGTCCGGGACCCCGGCGTGTCCGTTCCCGAGA 825  
Db 1294 CTTGACCTCAGGTCATCCACCTAGTACGTGATCTGTCTGTGTGCGACGCTCCGAT 1353  
QY 826 ACGCACAAGGAGATCTCATGATTGACAGTCGATTTGATGGGAGCATGAGACAGA 885  
Db 1354 CCGCTCGGCTACGCGCTCATCCGGAAGCCTACAGGTGTGGGAGACCGCATCAAG 1413  
QY 886 ACTCA 891  
Db 1414 AGCCA 1419

RESULT 10  
US-09-040-984-63  
Sequence 63, Application US/09040984  
Patent No. 6210883  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Wang, Tongtong  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS  
TITLE OF INVENTION: OF LUNG CANCER  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,984  
FILING DATE: 18-MAR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.456  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-282-6031  
TELEX:  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-040-984-63

Query March 3.1%; Score 62.6; DB 4; Length 731;  
Best Local Similarity 58.2%; Pred. No. 1.9e-08;  
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGTGCTCAAGAGAGTGGCTTTCACCTGACCTCCCTCTACCCGCTGTAAGAAC 761  
Db 3 TAGTCATAAAGGAGTGGCTTTCACAGTGGCGGCTTGGGCGCACCTGCTGGAGACC 62  
QY 762 CTTCCCTCAGCTCATATGCTGACCTGGTCCGGGACCCCGGCGGTGTCCGTTCC 821  
Db 63 CGGCCCTGAGACCTCAAGTATCCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 122

QY 822 GAGACGACACAAGAGATCTCATGATTGACAGTGCATGTGATGGGACCATGAGC 881  
Db 123 GATCCCTGCGCCACCGGCTCATCCGTGAGAGCTTACAGGTGTGGGACCGGAGACC 182  
QY 882 AGAAGCTCA 890  
Db 183 GCGAGCTCA 191

RESULT 11  
US-09-123-912-63  
Sequence 63, Application US/09123912A  
Patent No. 6312695  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Wang, Tongtong  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER  
FILE REFERENCE: 210121.455C1  
CURRENT APPLICATION NUMBER: US/09/123,912A  
CURRENT FILING DATE: 1998-07-27  
PRIOR FILING DATE: 09/040,802  
SOFTWARE: PatentIn Ver. 2.0  
NUMBER OF SEQ ID NOS: 114  
SEQ ID NO 63  
LENGTH: 731  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (236)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (249)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (263)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (288)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (312)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (317)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (323)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (326)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (337)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (352)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (362)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (370)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (376)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (400)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base





Db 1231 TCATCCACTGTTGTCCTGATATCCCGCCGGTGGCGAGTTACAGGATCCGCTCGCCACG 1290  
QY 840 ATCTCATGATGACAGTGGCATTTGTGATG----- 868  
Db 1291 GCTCATCCGCTGAGAGCTACAGTGTGTGGCAGACCCGAGACCCGAGCTACCGCATGC 1350  
QY 869 -----GGGACGATATGACGAAGAACCTCAAGAGAGG----- 899  
Db 1351 CCTCTTGAGAGCGCGGGCCACAGCTTGGCCGCAAGAGAGAGGCGCTGGGCGGCCCG 1410  
QY 900 -----ACCAACCTACTATGATGATGAGGATCATCTGCCAAAGCCAGCTGAGATTCACA 953  
Db 1411 CAGACTACAGGCTCTGGGCGCTATGAGGATCATCTGCATATGATATGCTAAGACGCTGC 1470  
QY 954 AGACATTCAGATCTTGTGCCAAGGCGCTGCAGAGACCTTCCTGCTTGTGCGCTATAGAG 1013  
Db 1471 AGAGAGCCCTGACAGCCCGCTGACGTGTGAGAGGCGCACCTGTGTGCTCGGATACAGG 1530  
QY 1014 ACCTGGCTGAGCCCTGTGTGGCCAGACTTCCGCAATGTATGATTTGTTGGATTTGAT 1073  
Db 1531 ACCTGTGGAGAGACCCGCTCAAGACACTAGGAGAGATGTACGATTTTGTGGACTGTGG 1590  
QY 1074 TCTTGCCCATCTTCAACACTGGGTGATTAACATCACCGAGCGAGCGCATGGGTGACC 1133  
Db 1591 TGAGCCCGGAATGAGACAGATTGCTCCTGAACATGACAGTGGCTCGGCTCCTCTCA 1650  
QY 1134 AGCTTTTCACACAAATGCGAGGATGCGCTATATGTCCTCCAGGCTTGGCGCTGTCT 1193  
Db 1651 AGCTTTGTGTGATCTGACAGCATATGCGACGACGCGCCCAATGCTGGCGGACCGCC 1710  
QY 1194 TGCCCTATGAAGAAGTTCTGACCTTCAGAAAGCCTGTGGCGATGCGCATGATTTGCTG 1253  
Db 1711 TCACCTTCACAGATCAACAGAGTGGAGAGTGTTCCTACAGCCCATGCGCTCTG 1770  
QY 1254 GCTACCGCACGTCA 1268  
Db 1771 GCTATGAGCGGTCA 1785

RESULT 7  
US-09-471-867-3  
Sequence 3, Application US/09471867  
Patent No. 6435289  
GENERAL INFORMATION:  
APPLICANT: Uchimura, Kenji  
APPLICANT: Muramatsu, Hideki  
APPLICANT: Kadamatsu, Kenji  
APPLICANT: Kanuchi, Reiji  
APPLICANT: Habuchi, Osami  
APPLICANT: Muramatsu, Takashi  
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND  
FILE REFERENCE: TOYAMA1.001AUS  
CURRENT APPLICATION NUMBER: US/09/471.867  
CURRENT FILING DATE: 1999-12-23  
EARLIER APPLICATION NUMBER: US 09/263.023  
EARLIER FILING DATE: 1999-03-05  
EARLIER APPLICATION NUMBER: JP 10-54007  
EARLIER FILING DATE: 1998-03-05  
EARLIER APPLICATION NUMBER: JP 10-177844  
EARLIER FILING DATE: 1998-06-24  
NUMBER OF SEQ. ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 2409  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3390)...(1841)  
US-09-471-867-3

Query Match 5.9%; Score 119.4; DB 4; Length 2409;  
Best Local Similarity 49.1%; Pred. No. 1e-24;  
Matches 508; Conservative 0; Mismatches 446; Indels 81; Gaps 4;  
QY 315 TGCTGTTCTGCTTCTCCGCGCTGTGCGCTCTTCTTTGTGGGAGCTTTTGGGACG 374  
Db 751 TGTAGCTTTACACAGATGGCGCTGTGCTGTCTGCTCTTCTGGGAGCTATTCACACAGA 810  
QY 375 ACCAGATGTTTCTACCTGATGAGACCCGCTGGCAGCTGTGATGACCTTCAAGCAGA 434  
Db 811 ATCCGAGGTGTCTTCTCTACAGCCAGTGTGCTATGAGCAAAACGTATCCGG 870  
QY 435 GCACGCGCTGAGATGCGCATGTGCTGTGCGGATCTGATAGGCGCGCTCTTGTGCG 494  
Db 871 GAGAGCGCTTCCCTGCGAGGGGAGCGCGGACATGCTGAGCCCTTTACCCCTGCG 930  
QY 495 ACATGAGCTCTTGTGATGCTTACATGAACTGTGTCCCGGAGAC-----AGT 542  
Db 931 ACCTCTCTGTCTCCAGTTGTATAGCCCGCGGAGCGGGGGGCGCAACTCACACAGC 990  
QY 543 CCAGCCTCTTTCATGTTGGGAGACAGCGGCGCTGTGTTCTGACCTGCTGTGACATCA 602  
Db 991 TGGGCACTTTCGCGCGCACACACAGAGTGTGTGCTGTGACACTTGTGCGCCGCT 1050  
QY 603 TCCACAGATGAATAATCATCCCGGCTCACTGCAAGGCTC---CTGTGCACTCAACAGC 659  
Db 1051 ACAGCAAGAGAGTGTGTGGGTGTGTGAGACACCGGTGTGCAAGAAGTCCCGCACAGC 1110  
QY 660 CTTTAAAGTGTGTGAGAGGCTGTGCTCTTACAGCAGCTGTGTCAAGAGGTGC 719  
Db 1111 GCTGTGCGGTTTGTGAGGAGGTGCGCAATGATACGCACTATCATAAAGGTTGTGC 1170  
QY 720 GCTTCTTCAACCTGACGTCTCTTACCCGCTGTGAAAGCCCTCTTCAACCTGACATA 779  
Db 1171 GCTCTTCACAGTGGCGGTCTTGGCGCCACTGTCTGAGAACCCGCGCTGAGCTCAAGG 1230  
QY 780 TGTGCACTGTGTGCGGAGACCCCGGCGCTTTCCTGCTTCCGAGAACGACCAAGGAG 839  
Db 1231 TCATCCACTGTTGCTGTATCCCGCGGTGCGGAGTTACAGGATTCGCTGCGCCACG 1290  
QY 840 ATCTCATGATGACAGTTCGATTTGTGATG----- 868  
Db 1291 GCTCATCCGTGAGACCTTACAGGTGTGTGCGAGCCGAGACCCGAGCTCACCGCATGC 1350  
QY 869 -----GGGACGATGAGCAGAACTCAAGAGAGG----- 899  
Db 1351 CCTTTTGAGCGCGCGGCCACAGCTTGGCGCAAGAGAGGAGCGGTGGCGGCCCG 1410  
QY 900 -----ACCAACCTACTATGTGATGACAGTTCGCAAGCCAGCTGAGATCTACA 953  
Db 1411 CAGACTTACACGCTCTGGCGCTATGAGAGTCTGCAATGATATGCTAAGACCTGTC 1470  
QY 954 AGACATTCAGTCTTGTGCCAAGCGCCCTGCAGAGACGCTACCTGCTGTGCGCTATGAG 1013  
Db 1471 AGACAGCCCTGAGAGCCCGCTGACTGAGCTGACGAGGCACTACTGTGTGGCGCTGAGG 1530  
QY 1014 ACCTGCTGTGAGCCCTGTGTGGCCAGACTTCCGATGTATGATATTCGTGGATTTGAT 1073  
Db 1531 ACCTGTGGAGAACCCCGCAAGACATGAGAGGTGTACATTTTGTGGAGCTGTGC 1590  
QY 1074 TCTTGCCCATCTTCAAGACTGCTGATTAACATACCGAGCGAGGACATGGGTGACC 1133  
Db 1591 TGAGCCCGGAATGAGAGGTTTGGCTGTGACATGACAGTGTGCTCGGCTCTCTCA 1650  
QY 1134 AGCTTTTCACACAAATGCGAGGATGCTTAAATGTCTCCAGGCTTGGCGCTGTCTT 1193  
Db 1651 AGCTTTGTGTGATCTGTGACCAATGCCACGAGCGCGCAATGCTGGCGGAGACCGCC 1710  
QY 1194 TGCCCTATGAAGAAGTTTGTGACTTCAAGAAAGCGTGTGGGATGCCATGATTTGCTG 1253  
Db 1711 TCACCTTCACAGATCAACAGTGTGAGAGTGTTCCTACAGCCCATGCGCTCTG 1770  
QY 1254 GCTACCGCACGTCA 1268



DB 488 GACCTCTGCGAGACCTCTAGACTGCGACCTCTACTTCTGAGAGAACTACATCAAGCCG 547  
QY 527 GGTCCCGGAGAGAGTCCAGCC-----TCTTTCAGTGGGAGAGACAGCCGGCCCTGTGT 580  
DB 548 CCGCCGCTCAACACACACACACAGAGATCTTCCGCGCGGGGCGACCGGGTCTCTGCG 607  
QY 581 TCTGCACTGCTGTGACATCATCTCCACAAGAAATCAATCC--CCCGGGCTCACTGC 637  
DB 608 TCCCGGCTGTGTGACACCTCCGCGGGCCAGCGACCTGTGCTCTGAGAGAGGGGACTGT 667  
QY 638 AGGCTCTGTGCACTCAAGACCCCTTGTAGTGTGAGAGAGCCCTGCGCTCTCTACAG 697  
DB 668 GTGCGCAAGTGTGCGTACTCAACCTCAACCTGCGCGCGGCGGCTGTGCGAGCGTAGC 727  
QY 698 CACGTGTGCTCAAGAGAGTGTGCTTCTTCACTGACGTCTCTACCTCCCTGCTGAAA 757  
DB 728 CACGTGTGCTCAAGAGAGTGTGCTTCTTCACTGACGTCTCTACCTCCCTGCTGAAA 787  
QY 758 GACCCCTCTCTCAACCTGCAATGTGTGCTGCTGCGGAGCCCGGGGCTGTTCGT 817  
DB 788 GACCCGCGATTAACCTCAAGGTCTACCTGCTGCTGCGAGCCCGCGGCTATCTGCT 847  
QY 818 TCCCGAGAGCGCAAGAGAGATCTGATGATGACAGTGTGCAATGTGTGAGGGAGCAT 877  
DB 848 TCGGCGAGCGAGACCTTCCGAGACGTACCGGCTGTGGGCGCTGTGTACGGCAGCGG 907  
QY 878 GAGCAGAACTCAAGAGAGAGAGACCACTACTATGTGATGAGAGTCACTGTCCAAAGC 937  
DB 908 AGGAACCTCAACACCTGACGTGACGACCTGACCAAGGTGTGCGAGGACTCTCTCAAC 967  
QY 938 CAGCTGAGATCTCAAGACCATCTCTGCTGCTGCGCAAGGCGCTGCGAGAGCGTCTG 997  
DB 968 TCGGTGTGCTGCGGCTCTGATGCGGCGCGCGGCTGCAAGGCGAG-----TTCATG 1018  
QY 998 CTTGTGCGTATGAGAGACTGTGCTGAGCCCTGTGCGCCGAGACTTCCGAGATGTAA 1057  
DB 1019 TTGTTGCGCTACAGAGACGTGCGCTGCGAGCCCTATGAAGAACCGAGATCTACGG 1078  
QY 1058 TTCGTGGGATTTGATCTCTGCCCCATCTTCAAGCTGSGTGATATACATCACCAGGC 1117  
DB 1079 TTCTGTGGCATCCCGCTGAGACAGACGTGCGCGCTGTGATCCAGAACACACCGGGG 1138  
QY 1118 AAGGCGATGGGTGACACAGCTTTCACACAAATGCCAGGATCCCTTAATGTCTCCAG 1177  
DB 1139 GACCCCAACCCGCGGAGAGCAACAAATACGGACCGGTGCGAATCGGCGGCGAGCGG 1198  
QY 1178 GCTTGGCGCTGTGCTTTGCCCCATGAAGAGTTTCTGACTTCAGAAAGCTGTGGCGAT 1237  
DB 1199 AAGTGGCGCTTCCGCTCTCTCTACAGATGCTGCGCTTTGCGCAGAAAGCGCTCCAG 1258  
QY 1238 GCCATGATTTGCTGGGCTACCGGCGACGTGAGATCTGAACAAGAGAGAAAGCTGTG 1297  
DB 1259 GTGCTGCGCCAGCTGCGCTCTCTCTACAGATGCTGCGCTTTGCGCAGAAAGCGCT 1318  
QY 1298 CTGATCTTCTG 1309  
DB 1319 GTGACCTGTG 1330

## RESULT 5

US-08-655-878-1  
; Sequence 1, Application US/08555878  
; Patent No. 5827713  
; GENERAL INFORMATION:  
; APPLICANT: FUKUTA, MASAKAZU  
; APPLICANT: HANBUCHI, OSAMI  
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESS:  
; STREET:  
; CITY:

STATE:  
COUNTRY:  
ZIP:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentId  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/655,878  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2354  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Chick  
TISSUE TYPE: Embryo chondrocyte  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 211..1584  
FEATURE:  
IDENTIFICATION METHOD: P  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 211..309  
FEATURE:  
IDENTIFICATION METHOD: P  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 310..1584  
FEATURE:  
IDENTIFICATION METHOD: P  
FEATURE:  
NAME/KEY: transmembrane domain  
LOCATION: 280..321  
FEATURE:  
IDENTIFICATION METHOD: P  
FEATURE:  
NAME/KEY: potential N-glycosylation site  
LOCATION: 394..402  
FEATURE:  
IDENTIFICATION METHOD: S  
FEATURE:  
NAME/KEY: potential N-glycosylation site  
LOCATION: 427..435  
FEATURE:  
IDENTIFICATION METHOD: S  
FEATURE:  
NAME/KEY: potential N-glycosylation site  
LOCATION: 493..501  
FEATURE:  
IDENTIFICATION METHOD: S  
FEATURE:  
NAME/KEY: potential N-glycosylation site  
LOCATION: 916..924  
FEATURE:  
IDENTIFICATION METHOD: S  
FEATURE:  
NAME/KEY: potential N-glycosylation site  
LOCATION: 1405..1413  
FEATURE:  
IDENTIFICATION METHOD: S  
FEATURE:  
NAME/KEY: potential N-glycosylation site  
LOCATION: 1537..1545  
FEATURE:  
IDENTIFICATION METHOD: S

US-08-655-878-1

```

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
FILING DATE:
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOYAM21.001A05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2156
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Human
TISSUE TYPE: Fetal brain
FEATURE:
NAME/KEY: CDS
LOCATION: 147..1583
IDENTIFICATION METHOD: S
US-08-899-514-1

```

Query Match 7.0%; Score 143.2; DB 2; Length 2156;

Best Local Similarity 50.6%; Pred. No. 1.1e-31; Matches 489; Conservative 0; Mismatches 453; Indels 24; Gaps 5;

```

QY 311 CACGATGCTGTTCTGCTTCTTCTGCGCTCTGCTCTTCTTCTTGTGGGACAGCTTTTGGG 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 546 CACGTCGTCATAGTGGCCACGACGCGCGCTCTGTCGGGAGATTTCTTAC 605
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 CAGCACCAGATGTTTCTTACCTATGAGCCCGCTGGCAGGTGGA-----TGACC 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 606 CAGCGGGGCAACATCTTCTTACCTCTTCGAGCGCTGTGGACATGAGCGCAGAGTGTC 665
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 425 TTCAGCAGAG-----CACCGCTGATGCTGACATGAGCTGTGGGATGTATGCG 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 666 TTCGAGCGGGGGGCGCCAGCGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 725
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 GCGCTCTCTGTCGCGCATGAGGCTCTTGTAGTACCTACATGAACTGCTCCCGGAGA 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 726 CAGCTCTCTGTCGAGACCTGCTGAGCAGCATTCATCAGCGCTCCCGAGAGAC 785
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 539 CAGTCACACCTCTTCACT-----GGGAGAACAGCGCGGCTGTGTGACACTGCC 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 786 CACCTGATCTAGTTTATTTCCCGCGGGCTTCAGCGCTCCCTGTGGAGAGACCGCTC 845
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 593 TGTGACATCATCCACAAGATGAATCATCCCGGCTCACTGACAGGCTCTGTGAGT 652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 846 TGTAGCCCTTGTCAAGAGAGTCTTCAGAGTACACTGCAAGAACCGCGCTGGGCG 905
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 653 CACAGCCCTTTGAGTGTGGAGAGAGCCCTGCGCTCTTCAAGCCAGCTGTGCTCAAG 712
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 906 CCGCTCAAGTAGAGCTGCGCGAGAGCTGCGCGCAAGAGACATGCGCTCAAG 965
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 713 GAGTGTGCTCTTCAACCTGAGTCCCTTACCGCGTGTGGAAGACCCCTCCCTCAAC 772
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 966 GGGGTGGCATCCGGAGCTGAGTCTTGAAGCGCTGGCGGAGAGACCCCGCTGGAC 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 773 CTGCTATGCTGACAGTGTGCGGAGACCCCGGCGCTGTTCCTGCTTCCCGAAGACGACA 832
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 1026 CTGCGCTCATCCAGCTGTGCGCGAGACCCCGGCTGCTGCGCTGCGCATGCTGCGC 1085
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 833 AAGGAGATCTCATGATTGACAGCTGCACTGTGATGAGGAGCAGTGAAGAACTCAAG 892
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1086 TTCGCCGCAAGTATAGAGCTGGAAGAGTGTGTGACAGCAGAGGAGCGGCTG 1145
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 893 AAGGAGACCAACCTACTATGTATGATGAGTGTGATGAGTGTGATGAGTGTGATGAGT 952
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1146 AGGAGAGAGA---GGTGCAGCGCTGCGGGGCACTGCGAGACATTCGCTGTGCGG 1202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 953 AAGACATCCAGTCTTGGCCAAAGGCCCTGCAAGAGACGCTACCTGCTGTGCTATGAG 1012
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1203 GAGGTGGGCTGGGAGAGCCCGCTGCGTGGCGGCGCTACATGCTGTGGGCTTCAAG 1262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1013 GACCTGCTGAGCCCTGTGTCGCGAGACTTCCGATGATGATGATGATGATGATGATGAA 1072
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1263 GAGGTGACAGCGGCGCGCTGCAAGAGGCCCGGAGATGATCCGCTTCCCGCATCCCC 1322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1073 TTCTTGGCCCATCTTCAAGACCTGGTGTGATATCATGACCCGAGGAGCATGAGGTGAC 1132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1323 CTGATCCCGAGGTGGAAGACTGATCCAAAGAACAGCAGS---CGGCCACAGCGGC 1379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1133 CAGCTTTCACACAATGCGCAGGAGTGCCTTATGTCTCCAGGCTTGGCGCTGTCT 1192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1380 AGCGGCACTACTTCCAGCGAGAGAACTCTCGAGCAGTTCAGAGAGTGGCGCTTACAG 1439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1193 TTGCCCCATGAAGAAGTTTGTGACTTCAGAAAGCTGTGGGATGCTGATGATGCTG 1252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1440 ATGCGCTTCAAGCTGGCGCAGGTGTGAGGCCCCGCTGCGCTTCCATGCGCTTTC 1499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1253 GGCTAC 1258
    |||||
DB 1500 GGCTAC 1505
    |||||

```

RESULT 4

US-09-015-188-1

; Sequence 1, Application US/09015188C

; Patent No. 6399358

; GENERAL INFORMATION:

; APPLICANT: Tabas, Itra

; TITLE OF INVENTION: A Human Gene Encoding Human Chondrolectin

; FILE REFERENCE: JEFF-0231

; CURRENT APPLICATION NUMBER: US/09/015,188C

; NUMBER OF SEQ. ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.0

; SEQ. ID NO. 1

; LENGTH: 2190

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-015-188-1

Query Match 6.6%; Score 133.6; DB 4; Length 2190;

Best Local Similarity 49.0%; Pred. No. 7e-29; Matches 506; Conservative 0; Mismatches 484; Indels 42; Gaps 4;

```

QY 311 CAGTGTGCTTCTGCTTCTTCTGCGCTCTGCTCTTCTTCTTGTGGGAGCTTTTGGG 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 308 CACATCTCATCTGCGCACACAGCGAGGCGCTCTCTCTGCTGCGCACAGCTTCAAC 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 CAGCACCAGATGTTTCTTCACTGATGAGCCCGCTGAGCAGTGTGATGACCTTCAAG 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 368 CACACCTGAGAGCTTCTTCACTGATGAGCCCGCTTACACGCTCCAGACAGCAGCTATC 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 CAGAGCACCCTGAGATGCTGACATGAGCTGTGCGG-----466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 428 CCGCGCTTACACAGGCGCAGAGCCCGCGAGCGGCTGATGCTAGGCGCAGCCGC 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 467 GATCTGATACGGCGCTTCTTGTGAGACATGAGGCTTTGTGATGCTTACATGAGACCT 526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
OY 121 CAGCTTCCG6GAGGAGTGCCTTCTCAGGCCGCTTGCAGAGCTTCCACTTCAGAC 180
Db 121 CAGCTTCCG6GAGGAGTGCCTTCTCAGGCCGCTTGCAGAGCTTCCACTTCAGAC 180
OY 181 AATGCTACTGCTAAAAAATGAAGCTCCTCTCTTTCTGTTTCCAGATGSCCATCTT 240
Db 181 AATGCTACTGCTAAAAAATGAAGCTCCTCTCTTTCTGTTTCCAGATGSCCATCTT 240
OY 241 GGCTCTATTCTTCCAGATGTACAGCCACACATCAGCTCCCTGTCTATGAAGGACAGCC 300
Db 241 GGCTCTATTCTTCCAGATGTACAGCCACACATCAGCTCCCTGTCTATGAAGGACAGCC 300
OY 301 CGAGCGCATGACAGCTGTGTCTGTCTTCTGCGCTGCTGGCTCTTCTTTGAGGAGCA 360
Db 301 CGAGCGCATGACAGCTGTGTCTGTCTTCTGCGCTGCTGGCTCTTCTTTGAGGAGCA 360
OY 361 GCTTTTGGGAGACACCCAGATGTTTCTACCTGATGAGCCGCCCTGGCAGCTGTGGAT 420
Db 361 GCTTTTGGGAGACACCCAGATGTTTCTACCTGATGAGCCGCCCTGGCAGCTGTGGAT 420
OY 421 GACCTTCAAGCAGACACCGCTGGATGCTGACATGGCTGTGGGATCTGATACGAGC 480
Db 421 GACCTTCAAGCAGACACCGCTGGATGCTGACATGGCTGTGGGATCTGATACGAGC 480
OY 481 GGTCTTCTTGTGCGACATGAGCGTCTTTGATGCTTACATGGAACCTGTCCCGAGACA 540
Db 481 GGTCTTCTTGTGCGACATGAGCGTCTTTGATGCTTACATGGAACCTGTCCCGAGACA 540
OY 541 GTCCAGCCTCTTCAAGTGGAGAACAGCCGCTGTGTCTGCTGACCTGCTGTGACAT 600
Db 541 GTCCAGCCTCTTCAAGTGGAGAACAGCCGCTGTGTCTGCTGACCTGCTGTGACAT 600
OY 601 CATCCCAAGATGAATCATCCCGCGGCTCAGTGCAGAGGCTGCTGTGACATCAACAGC 660
Db 601 CATCCCAAGATGAATCATCCCGCGGCTCAGTGCAGAGGCTGCTGTGACATCAACAGC 660
OY 661 CTTTGAAGTGGTGGAGAGGCTGCTGCTTCTTACAGCCAGCTGTGTCTGAGAGGCTGG 720
Db 661 CTTTGAAGTGGTGGAGAGGCTGCTGCTTCTTACAGCCAGCTGTGTCTGAGAGGCTGG 720
OY 721 CTTTCTCAACCTGAGCTGCTCTACCCGCTGTGAAAGACCCCTCCCTCAACCTGCATAT 780
Db 721 CTTTCTCAACCTGAGCTGCTCTACCCGCTGTGAAAGACCCCTCCCTCAACCTGCATAT 780
OY 781 CGTGCACTGCTCGGAGACCCCGGCGCTGTCCCTTCCCGAGACGACACAAAGSAGA 840
Db 781 CGTGCACTGCTCGGAGACCCCGGCGCTGTCCCTTCCCGAGACGACACAAAGSAGA 840
OY 841 TCTCATGTGACAGTGCATTTGATGGGAGCAGATGAGAGAAACCAAGAGAGAGA 900
Db 841 TCTCATGTGACAGTGCATTTGATGGGAGCAGATGAGAGAAACCAAGAGAGAGA 900
OY 901 CCAACCCCTACTATGTATGATGAGTGTGATGAGAGAGCAGCTGAGATCTCAAGAGCAT 960
Db 901 CCAACCCCTACTATGTATGATGAGTGTGATGAGAGAGCAGCTGAGATCTCAAGAGCAT 960
OY 961 CCAAGTCTTGGCCCAAGGCTGAGAGAGAGCTTGTGCGCTATGAGAGCTGGC 1020
Db 961 CCAAGTCTTGGCCCAAGGCTGAGAGAGAGCTTGTGCGCTATGAGAGCTGGC 1020
OY 1021 TCGAGCCCTTGGCCCAAGCTTCCCAATGTATGAAATTCGTGGATTTGAAATCTTGGC 1080
Db 1021 TCGAGCCCTTGGCCCAAGCTTCCCAATGTATGAAATTCGTGGATTTGAAATCTTGGC 1080
OY 1081 CCAATTTAGAGCTGTGGTGCATTAACATACCCGAGGCAAGGGCATGGGTGACACGCTTT 1140
Db 1081 CCAATTTAGAGCTGTGGTGCATTAACATACCCGAGGCAAGGGCATGGGTGACACGCTTT 1140
OY 1141 CCACCAAAATGCCAGGAGTGCCTTATATGTCTCCCAAGGCTGTGGCTGTGGCTTGA 1200
Db 1141 CCACCAAAATGCCAGGAGTGCCTTATATGTCTCCCAAGGCTGTGGCTGTGGCTTGA 1200
OY 1201 TGAAGAGTTTCTGACTTCAGAAAGCCTGTGGGATGCCATGAATTTGCTGGGCTACG 1260
Db 1201 TGAAGAGTTTCTGACTTCAGAAAGCCTGTGGGATGCCATGAATTTGCTGGGCTACG 1260
```

---

```
Db 1201 TGAAGAGTTTCTGACTTCAGAAAGCCTGTGGGATGCCATGAATTTGCTGGGCTACG 1260
OY 1261 CCAGCTCAGATCTGAAACAGAGAGAACTGTGTTGATCTTCTCTTACCTGGAC 1320
Db 1261 CCAGCTCAGATCTGAAACAGAGAGAACTGTGTTGATCTTCTCTTACCTGGAC 1320
OY 1321 TGTCCCTGACCAATTCACATGAAGGTTGAGAGGCTTGTGCTGCCACCTGGGTGAGCC 1380
Db 1321 TGTCCCTGACCAATTCACATGAAGGTTGAGAGGCTTGTGCTGCCACCTGGGTGAGCC 1380
OY 1381 TCACTCACTTCTCTGATGCTTCTGAGCTTCCCTACATCTCTGAGCCTTAACATCAGT 1440
Db 1381 TCACTCACTTCTCTGATGCTTCTGAGCCTTCCCTACATCTCTGAGCCTTAACATCAGT 1440
OY 1441 TCTGTGGGTATTCACACTGAGTGTGATGTGTGTCCACAGTGTCTCAAGCAGAGAGCTTTT 1500
Db 1441 TCTGTGGGTATTCACACTGAGTGTGATGTGTGTCCACAGTGTCTCAAGCAGAGAGCTTTT 1500
OY 1501 GTGTCCATGCTTGTCTGTAAGAAACAGACTGGGAGACCTTATGTGAGCAGACATCCAC 1560
Db 1501 GTGTCCATGCTTGTGTGTAAGAAACAGACTGGGAGACCTTATGTGAGCAGACATCCAC 1560
OY 1561 CAGTGAACAGGCTATTTGCTCTTCTTCTTCTTCTGATCTTCTGCTGTGGGAGCTTGA 1620
Db 1561 CAGTGAACAGGCTATTTGCTCTTCTTCTTCTTCTGATCTTCTGCTGTGGGAGCTTGA 1620
OY 1621 AGACTTTGTGGCTGAGAGGCTATTGATTAAGCAGCAGACAGTATCAGTGAATGATCATTA 1680
Db 1621 AGACTTTGTGGCTGAGAGGCTATTGATTAAGCAGCAGACAGTATCAGTGAATGATCATTA 1680
OY 1681 ACCTCCCTGTCCACATCTTGGCCATGAGGAGATGATCTTTTCCACCAAGAGCTCACAC 1740
Db 1681 ACCTCCCTGTCCACATCTTGGCCATGAGGAGATGATCTTTTCCACCAAGAGCTCACAC 1740
OY 1741 ATTTTCCACAGATGATCAATCTGAGCCCTTGTGAGCTTCCAGTGAATTCAGAGAGGA 1800
Db 1741 ATTTTCCACAGATGATCAATCTGAGCCCTTGTGAGCTTCCAGTGAATTCAGAGAGGA 1800
OY 1801 GTGGGAACAAGTTGATGCTACTTATGAGTGTGACATCAGCTATCGTAAATCAGA 1860
Db 1801 GTGGGAACAAGTTGATGCTACTTATGAGTGTGACATCAGCTATCGTAAATCAGA 1860
OY 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTCTTAAGTTCAAGAGGTGCTGGGC 1920
Db 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTCTTAAGTTCAAGAGGTGCTGGGC 1920
OY 1921 TGCATTTGAATATCACTTCCCTCTGATTTTCCATCAATGAGAGCTTGAACCTGTG 1980
Db 1921 TGCATTTGAATATCACTTCCCTCTGATTTTCCATCAATGAGAGCTTGAACCTGTG 1980
OY 1981 AAGCTGCCATCTGTAACTAAATCCCAATTAAGAAAAA 2032
Db 1981 AAGCTGCCATCTGTAACTAAATCCCAATTAAGAAAAA 2032
```

RESULT 3  
US-08-899-514-1  
: Sequence 1, Application US/0889514  
: Patent No. 5910581  
: GENERAL INFORMATION:  
: APPLICANT: HABUCHI, OSAMI  
: APPLICANT: FUKUDA, MASAKAZU  
: TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN  
: TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP  
: STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR  
: CITY: NEWPORT BEACH  
: STATE: CALIFORNIA  
: COUNTRY: US

Db 361 GCTTTTGGGAGACACCCAGATGTTTCTACGTATGAGCCCGCTGGCAGCTGTGAT 420  
QY 421 GACCTTCAAGCAGACACCGCTGGATGCTGCACATGGCTGGGGAGTGTGATACGGCC 480  
Db 421 GACCTTCAAGCAGACACCGCTGGATGCTGCACATGGCTGGGGAGTGTGATACGGCC 480  
QY 481 GCTCTTCTTGTGCGACATGAGCGCTTTTGATGCCATGAGAACCTGGTCCCGAGACA 540  
Db 481 GCTCTTCTTGTGCGACATGAGCGCTTTTGATGCCATGAGAACCTGGTCCCGAGACA 540  
QY 541 GTCGAGCTCTTTCAGTGGAGAACAGCCGGCCCTGCTTCTGCACCTGGCTGTGACAT 600  
Db 541 GTCGAGCTCTTTCAGTGGAGAACAGCCGGCCCTGCTTCTGCACCTGGCTGTGACAT 600  
QY 601 CATCCACAGATGATAATCATCCCGCGGCTCAGTGCAGGCTCCTGTCAGTCAACAGCC 660  
Db 601 CATCCACAGATGATAATCATCCCGCGGCTCAGTGCAGGCTCCTGTCAGTCAACAGCC 660  
QY 661 CTTGAGGTGGTGGAGAACGCCCTGCGCTCTACAGCCAGCTGGTGGTCAAGAGGTGG 720  
Db 661 CTTGAGGTGGTGGAGAACGCCCTGCGCTCTACAGCCAGCTGGTGGTCAAGAGGTGG 720  
QY 721 CTTCTTCAACCTGAGTGCCTCTACCCGCTGTGAAAAGCCCTCCCTCAACCTGCATAT 780  
Db 721 CTTCTTCAACCTGAGTGCCTCTACCCGCTGTGAAAAGCCCTCCCTCAACCTGCATAT 780  
QY 781 CGTGACCTGGTCCGGGAGCCCGCGGGCGGTGTCGGTTCCGGAGACGACACAAAGGAGA 840  
Db 781 CGTGACCTGGTCCGGGAGCCCGCGGGCGGTGTCGGTTCCGGAGACGACACAAAGGAGA 840  
QY 841 TCTCATGATGACAGTGCATTTGATGGGGCAGCATGAGAGAACTCAAGAGAGAGA 900  
Db 841 TCTCATGATGACAGTGCATTTGATGGGGCAGCATGAGAGAACTCAAGAGAGAGA 900  
QY 901 CCAACCTGATATGATGACAGTGCATTTGATGGGGCAGCATGAGAGAACTCAAGAGAGAT 960  
Db 901 CCAACCTGATATGATGACAGTGCATTTGATGGGGCAGCATGAGAGAACTCAAGAGAGAT 960  
QY 961 CCAAGCTCTGGCCAAAGGGCCCTGCGAGAAAGCTTACCTGCTGTGGCTATGAGAGCTGGC 1020  
Db 961 CCAAGCTCTGGCCAAAGGGCCCTGCGAGAAAGCTTACCTGCTGTGGCTATGAGAGCTGGC 1020  
QY 1021 TCGAGCCCTGTGGCCAGACATTCGCAATGATGAATGCTGGGATGGAATCTTGCC 1080  
Db 1021 TCGAGCCCTGTGGCCAGACATTCGCAATGATGAATGCTGGGATGGAATCTTGCC 1080  
QY 1081 CCACTTTCAGACCTGGGTGCAATACATCACCCGAGGCAAGGGCATGGGTGACACAGCTTT 1140  
Db 1081 CCACTTTCAGACCTGGGTGCAATACATCACCCGAGGCAAGGGCATGGGTGACACAGCTTT 1140  
QY 1141 CCACACAATGCCAGGGATGCCCTTAATGTCCTCCAGGCTGGGGCTGCTTGGCCCTA 1200  
Db 1141 CCACACAATGCCAGGGATGCCCTTAATGTCCTCCAGGCTGGGGCTGCTTGGCCCTA 1200  
QY 1201 TGAAGAGGTTCTGACTTCAAGAAAGCCCTGTGGGATGCCATGATTTGCTGGGCTACG 1260  
Db 1201 TGAAGAGGTTCTGACTTCAAGAAAGCCCTGTGGGATGCCATGATTTGCTGGGCTACG 1260  
QY 1261 CCAGCTCAGATCTGAACAGAGACAGAAAGCTTGTGCTGAGTCTTGTCTACCTGGAC 1320  
Db 1261 CCAGCTCAGATCTGAACAGAGACAGAAAGCTTGTGCTGAGTCTTGTCTACCTGGAC 1320  
QY 1321 TGTCCCTGAGCAATTCACACTAGAGAGGTTGAGAAAGCTTTGCTGCCACCTGATGTACGCC 1380  
Db 1321 TGTCCCTGAGCAATTCACACTAGAGAGGTTGAGAAAGCTTTGCTGCCACCTGATGTACGCC 1380  
QY 1381 TCACTTCTTCTGATGATGCTTGTGAGCTTGTGCTGATCTGTGAGCTTCACTACATG 1440  
Db 1381 TCACTTCTTCTGATGATGCTTGTGAGCTTGTGCTGATCTGTGAGCTTCACTACATG 1440  
QY 1441 TCTGTGGGTATCACACTGAGTGTGAGTGTGTGCACACGTGTCAAGCAGAGAGACTTTT 1500  
Db 1441 TCTGTGGGTATCACACTGAGTGTGAGTGTGTGCACACGTGTCAAGCAGAGAGACTTTT 1500

QY 1501 GTGTCACATGCTGTGCTAGAAAAACAGACTGGGGAACTTATGTGACGACAGATCCAC 1560  
Db 1501 GTGTCACATGCTGTGCTAGAAAAACAGACTGGGGAACTTATGTGACGACAGATCCAC 1560  
QY 1561 CAGTGAACAGAGGATATTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1620  
Db 1561 CAGTGAACAGAGGATATTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1620  
QY 1621 AGACTTGTGGCTGGAGGCTTATTAAGCAGCAGACAGTATCAGTGAATTCATTA 1680  
Db 1621 AGACTTGTGGCTGGAGGCTTATTAAGCAGCAGACAGTATCAGTGAATTCATTA 1680  
QY 1681 ACCCTCCCTGACACATCTTCCCAATGGGCAATGATCTTTCACCAAGAGCTCACAG 1740  
Db 1681 ACCCTCCCTGACACATCTTCCCAATGGGCAATGATCTTTCACCAAGAGCTCACAG 1740  
QY 1741 ATTTCCACAGAGATGCAAAATTCAGCCCTTGGAGTTCCCACTGATTCAGAGAGAA 1800  
Db 1741 ATTTCCACAGAGATGCAAAATTCAGCCCTTGGAGTTCCCACTGATTCAGAGAGAA 1800  
QY 1801 GTGGGAACAAGTGGATGCTACTATGAGCTTGAACATACAGCTATCCGTAATCAGA 1860  
Db 1801 GTGGGAACAAGTGGATGCTACTATGAGCTTGAACATACAGCTATCCGTAATCAGA 1860  
QY 1861 AATATGAACAAATCTCTGACAAAAGCAAGCTCTTAAGTTCAAGAGGCTGGGC 1920  
Db 1861 AATATGAACAAATCTCTGACAAAAGCAAGCTCTTAAGTTCAAGAGGCTGGGC 1920  
QY 1921 TGCATTTGAATACACTTCCCTCTGCTGATTTTCCCATCAGATGAAGACTTTGACTGTG 1980  
Db 1921 TGCATTTGAATACACTTCCCTCTGCTGATTTTCCCATCAGATGAAGACTTTGACTGTG 1980  
QY 1981 AAGCTGCATCTGTATTAATAATCCCAATGAAGAAAAA 2032  
Db 1981 AAGCTGCATCTGTATTAATAATCCCAATGAAGAAAAA 2032

RESULT 2  
US-09-190-911-2  
; Sequence 2, Application US/09190911  
; Patent No. 6365365  
; GENERAL INFORMATION:  
; APPLICANT: Bistup, Annette  
; APPLICANT: Rosen, Steven D.  
; APPLICANT: Tangemann, Kirsten  
; APPLICANT: Hemmerlich, Stefan  
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
; FILE REFERENCE: 6510-107CIP  
; CURRENT APPLICATION NUMBER: US/09/190,911  
; EARLIER FILING DATE: 1998-11-12  
; EARLIER APPLICATION NUMBER: 09/045,284  
; EARLIER FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 2  
; LENGTH: 2032  
; TYPE: DNA  
; ORGANISM: H. sapiens  
US-09-190-911-2

Query Match 100.0%; Score 2032; DB 4; Length 2032;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCTCAGGCGCAGATGCTTCCAGTCTGGGGAAATGCTTCTCATTTGCTTCCAG 60  
Db 1 GGTCTCAGGCGCAGATGCTTCCAGTCTGGGGAAATGCTTCTCATTTGCTTCCAG 60  
QY 61 CCCACTCAAGAGATCTCCCAACCCCTTGAAGTCTGAGAGTCTTAAAGCTGTACTTCA 120  
Db 61 CCCACTCAAGAGATCTCCCAACCCCTTGAAGTCTGAGAGTCTTAAAGCTGTACTTCA 120

GenCore version 5.1.3  
Copyright (c) 1993 - 2003, CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:38:14 : Search time 80 Seconds

(without alignments)  
7789.390 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032

Sequence: 1 ggcctcgagccagcagcagcagcct.....atagaagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_5/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTBUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2032	100.0	2032	US-09-045-284A-1
2	2032	100.0	2032	US-09-190-911-2
3	143.2	7.0	2156	US-08-889-514-1
4	133.6	6.6	2190	US-09-015-188-1
5	130	6.4	2334	US-08-655-878-1
6	119.4	5.9	2409	US-09-263-023-3
7	119.4	5.9	2409	US-09-471-867-3
8	106.4	5.2	2150	US-09-263-023-1
9	106.4	5.2	2150	US-09-471-867-1
10	62.6	3.1	731	US-09-040-984-63
11	62.6	3.1	731	US-09-123-912-63
12	62.6	3.1	731	US-09-643-597-63
13	56.4	2.8	7218	US-08-232-463-14
14	37.6	1.9	3010	US-09-199-637A-168
15	37.2	1.8	3865	US-09-149-476-296
16	37.2	1.8	5894	US-08-665-259-24
17	37.2	1.8	5894	US-08-762-500-74
18	37.2	1.8	6525	US-08-762-500-74
19	37	1.8	37	US-09-045-284A-7
20	37	1.8	37	US-09-190-911-7
21	37	1.8	1965	US-09-178-252-26
22	36.8	1.8	2485	US-08-424-424B-1
23	36.8	1.8	2486	PCT-US94-05363A-1
24	36.4	1.8	44377	US-08-804-227C-7
25	36.4	1.8	44377	US-08-804-198-1
26	36	1.8	1003	US-09-149-476-162
27	36	1.8	5769	US-08-652-971-1

c	28	36	1.8	5769	2	US-08-991-258A-1	Sequence 1, Appl
c	29	36	1.8	5769	2	US-08-769-399-1	Sequence 1, Appl
c	30	36	1.8	5769	3	US-08-991-953A-1	Sequence 1, Appl
c	31	35.4	1.7	1882	3	US-09-096-398-1	Sequence 1, Appl
	32	35.4	1.7	1886	2	US-08-648-657-1	Sequence 1, Appl
	33	35.4	1.7	1689	2	US-08-648-657-2	Sequence 2, Appl
	34	35.4	1.7	1696	3	US-09-096-399-3	Sequence 3, Appl
	35	35.4	1.7	2499	1	US-07-977-434-1	Sequence 1, Appl
	36	35.4	1.7	2499	1	US-08-458-813-1	Sequence 1, Appl
	37	35.4	1.7	2499	3	US-07-602-848E-1	Sequence 1, Appl
	38	35.4	1.7	2499	4	US-09-587-856-1	Sequence 1, Appl
	39	35.4	1.7	2499	4	US-09-777-537-1	Sequence 1, Appl
	40	35.4	1.7	2499	4	US-09-777-537-1	Sequence 1, Appl
	41	35.4	1.7	2499	5	PCT-US91-07035-1	Sequence 1, Appl
	42	35.4	1.7	2505	2	US-08-757-653-167	Sequence 167, App
	43	35.4	1.7	2505	2	US-08-823-516-65	Sequence 65, App
	44	35.4	1.7	2505	3	US-08-758-038-106	Sequence 106, App
	45	35.4	1.7	2505	3	US-08-758-314-106	Sequence 106, App

## ALIGNMENTS

RESULT 1  
US-09-045-284A-1  
; Sequence 1, Application US/09045284A  
; Patent No. 6265192  
; GENERAL INFORMATION:  
; APPLICANT: Bistrup, Annette  
; APPLICANT: Rosen, Steven D.  
; APPLICANT: Hammerlich, Stefan  
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3  
; FILE REFERENCE: 6510-107051  
; CURRENT APPLICATION NUMBER: US/09/045,284A  
; CURRENT FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2032  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-045-284A-1

Query Match	100.0%	Score 2032:	DB 4:	Length 2032:
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 2032:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	1	GGCTCGAGGCGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	60	
DB	1	GGCTCGAGGCGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	60	
QY	61	CCCACTGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	120	
DB	61	CCCACTGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	120	
QY	121	CAGTCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	180	
DB	121	CAGTCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	180	
QY	181	AATGCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	240	
DB	181	AATGCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	240	
QY	241	GGCTCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	300	
DB	241	GGCTCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	300	
QY	301	CGAGGCGATGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	360	
DB	301	CGAGGCGATGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	360	
QY	361	GGCTCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	420	
DB	361	GGCTCTGAGGAGGAGGCTCTGAGTGGGGGAAATGCTTCCATTTGCTTCAG	420	



R. Emmett-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbip/image/image.html](http://www.bio.llnl.gov/dbip/image/image.html)

```
Search completed: January 10, 2003, 21:29:53
Job time : 2924 secs
```

```
Db 240 GAAGATCATCTGCAAAAGCCAGGTGACATAGTCAAGCCATCCAAACCCCTCCCTGAAGC 299
Qy 979 CCTGACGAGACGCTACTCTGTTGCGCTATAGAGACCTGGCTCCGAGCCCTGTGGCCCA 1038
Db 300 TTGCGACGACGCTACTCTGTTGCGCTATAGAGACCTGGCTCCGAGCCCTGTGGCCCA 359
Qy 1039 GACTTCCGAGATGATGAATCTGTGGATTTGAATCTTGGCCCATCTCAGACCTGGGT 1098
Db 360 GAGGACAGAGCTATTAATTT-GTGGGTTGATTTTGGCCCATCTCAGACCTGGGT 418
Qy 1099 GCATTAACATACCCGAGGAGGCGATGGGTGACACGCTTTCACACAAATGCCAGGA 1158
Db 419 TCACATATGTCACCCGGGGAAGGCGATGGGTGACACGCTTTCACACAAATGCCAGGA 478
Qy 1159 TGGCCTTAATGTTCCAGGCTTGGCGCTGATGCTTGGCCCTA-TGAAGGTTTTCGAC 1217
Db 479 CGCCCTCAAGCTCTCAGGCGGTGGCTGTGCTTACTTACCGAAGAGTTTCCGAC 538
Qy 1218 TTCAAGAAAGCTGTGGCGATGCCAT-GAATTGCTGGGCTACCGCCAGCTCAGATCTGAA 1276
Db 539 TTCAAGATGCTCGGTGAGGCTGATGATTTTGTGGTACCTCCAGGCTCAGATCTCAA 598
Qy 1277 CAGAAGACGAGAACCTGTTGCTGATCTTGTCTACCTGGACCTCCCTGAGCAATC 1336
Db 599 CAGAGACA-AGGCAACTGTCCTGCGATCTTCTCTCTCCATATCTTGGGGCAGCTC 657
Qy 1337 CACTAAG-AGGTTGAGAAAGCTTTGCTGCGACACTGGTGTCAACC 1380
Db 658 TTCCGAGACAGGTTAAGAGAGGCTGTGACACCCCTGTTCCAGGCC 702

RESULT 13
AM572390 419 bp mRNA linear EST 13-MAR-2000
LOCUS xu09a06.x2 NCI_CGAP_Col4 Homo sapiens cDNA clone IMAGE:2799634 3'
DEFINITION mRNA sequence.
ACCESSION AM572390
VERSION AM572390.1 GI:7237123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 419)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/nclogap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml
Seq. primer: 40UP from Gibco
High quality sequence stop: 343.
Location/Qualifiers
1. 419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2799634"
/clone_lib="NCI_CGAP_Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT 122 a 78 c 89 g 130 t
```

## ORIGIN

Query Match 16.4%, Score 332.4, DB 10; Length 419;  
Best Local Similarity 99.7%, Pred. No. 1.4e-60;  
Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 1684 TCCCGTCACATCTTGGCCCATGAGGATGATCTTTCACCAAGAGCTCACCAGCAT 1743
Db 419 TCCCGTCACATCTTGGCCCATGAGGATGATCTTTCACCAAGAGCTCACCAGCAT 360
Qy 1744 TTCCACAGAGATGCATATCTTGAGCCCTTGAGGTTCCAGTGATTCAGGAAGGAAGTG 1803
Db 359 TTCCACAGAGATGCATATCTTGAGCCCTTGAGGTTCCAGTGATTCAGGAAGGAAGTG 300
Qy 1804 GGAACAGGTTGAGATGCTTACTATGATCTTACCATCAGATCAGATCGATAGAAAT 1863
Db 239 GGAACAGGTTGAGATGCTTACTATGATCTTACCATCAGATCAGATCGATAGAAAT 240
Qy 1864 ATGAACAAATCTCTGACAAAGAGCAGCTCTTAAGTTTCACAGGGTGCCTGGCTGC 1923
Db 239 ATGAACAAATCTCTGACAAAGAGCAGCTCTTAAGTTTCACAGGGTGCCTGGCTGC 180
Qy 1924 ATTGAATATCATCTCCCTGCGATTTTCCATCACAATGAAGACTTTCCTGTGAAG 1983
Db 179 ATTGAATATCATCTCCCTGCGATTTTCCATCACAATGAAGACTTTCCTGTGAAG 120
Qy 1984 CTGCCATCTGTTAATCTAAATTTCCCAATTAAG 2017
Db 119 CTGCCATCTGTTAATCTAAATTTCCCAATTAAG 86

RESULT 14
B0947021 902 bp mRNA linear EST 21-AUG-2002
LOCUS B0947021
DEFINITION AGENCOURT 8951359 NCI_CGAP_C024 Mus musculus cDNA clone
IMAGE:6474928 5', mRNA sequence.
ACCESSION B0947021
VERSION B0947021.1 GI:22362499
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathu; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM14012 row: n column: 17
High quality sequence stop: 633.
Location/Qualifiers
1. 902
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6474928"
/clone_lib="NCI_CGAP_C024"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 228 a 246 c 210 g 213 t 3 others
ORIGIN
```



QY	362	CTTTTGGGAGACACCCAGATGTTTCTTACCTGATGGAGGCCGCCGGACGCTGGAGT	421
Db	244	CTTTTGGGAGACACCCAGATGTTTCTTACCTGATGGAGGCCGCCGGACGCTGGAGT <td>303</td>	303
QY	422	ACCTTCAAGACAGACGACCGCTGATGTGTCGCACATAGCTGTGCGGAGATCTGATACGGCC	481
Db	304	ACTTTCACACAGACACAGCTGGAAGGTGCACATAGCTGTGTCGGAGATCTTGTGCTTC	363
QY	482	GCTTCTCTGTGCGACATATAGCGCTCTTTGATGGCTACATAGGAACCTGTGTCCTGGAGACAG	541
Db	364	GCTTCTCTGTGCGACATATAGCGCTCTTTGATGGCTACATAGGAACCTGTGTCCTGGAGACAG	423
QY	542	TCCAGCCTCTTTCATGGGAGAGAACAGCCGGGCGCTGTGTTGTCACATGCTGTATACATC	601
Db	424	TCCAGCCTCTTTCATGGGAGAGAACAGCCGGGCGCTGTGTTGTCACATGCTGTATACATC	483
QY	602	ATCCCAAGATGAATATATATCCCGGGGCTCACTGCAGGCTCTCTGTGCATGACAGCC	661
Db	484	TTCCCTGCCCCAGATATAGCTACACCCCAAGCCTCAAGCTGTGCTGTGCGATGACAGCC	543
QY	662	TTTGAGTGTGTGAGAACAGCCGCGGCTGTACACCCAGCTGTGCTTAA-AGAGTGTG	720
Db	544	TTTGATATGTGTGAGAACAGCCGCGGCTGTACAGGCTCTGTGTACTACAGGAGTGTG	603
QY	721	CTTCTTACCTGCACTGCTCTTACCGCTGTGCTGTAAGAACCCCTCCCTCAACCTG-CAT	778
Db	604	TATTCCTACGCTGACGAGCCCTCTATGCCACTGTCTACGAGACCTTCCCTCAACCTGCA	663
QY	779	ATTCGACACCTGTGTGCGGAGAACCCCGGGGCGCTGTGCTGTGCGAGAACGCAAGGTA	838
Db	664	GTCGTGACCTGTGTGCGGAGAACCCCGGGGCGCTGTGCTGTGCGAGAACGCAAGGTA	722
QY	839	GATCTCATGATTGA-CAGTGTGATGTGATGAGGCGAGCATGACA 882	
Db	723	GAACTGTGTGTGACCACTGATCTGTGTGAGGAGCATTTGAA 767	

RESULT 10

BG964671

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

695 bp mRNA linear EST 12-JUN-2001

602813875f1 NCI\_CGAP\_C024 Mus musculus cDNA clone IMAGE:4986258 5'

mRNA sequence.

BG964671

BG964671.1 GI:14352308

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 695)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: scapbs-remail.nih.gov

Tissue Procurement: Jeffrey F. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

plate: LLM10995 row: f column: 19

High quality sequence stop: 684.

BASE COUNT  
ORIGIN

```

/note=Organ: colon; Vector: pCMV-Sport6; site_1: NotI;
site_2: SalI; cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. constructed by Life
Technologies. Note: this is a NCI-CGAP Library."
315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

```

Query Match	18.1%;	Score 368.6;	DB 13;	Length 695;
Best Local Similarity	74.9%;	Pred. No. 2.4e-68;		
Matches 501;	Conservative	0;	Mismatches 164;	Indels 4;
				Gaps 3







QY 1676 CATAAACCCTCCCTGTCACATCTTCCCAATGGGGAATGATCTTTTCACCAAAAGACTCA 1735  
 |||  
 Db 333 CATAAACCTCCCTGTCACATCTTCCCAATGGGGAATGATCTTTTCACCAAAAGACTCA 294  
 |||  
 QY 1736 CCAGCATTTTCCACAGAGATGCAAAATTTGAGCCCTTGGAGTTCCCAAGTGGATTCAAGA 1795  
 |||  
 Db 293 CCAGCATTTTCCACAGAGATGCAAAATTTGAGCCCTTGGAGTTCCCAAGTGGATTCAAGA 234  
 |||  
 QY 1796 AGGAAGTGGGAACAAGTGGATGCTTATATAGCTTGACATCAAGCTATGCTAA 1855  
 |||  
 Db 233 AGGAAGTGGGAACAAGTGGATGCTTATATAGCTTGACATCAAGCTATGCTAA 174  
 |||  
 QY 1856 TCAGAAATATGAACAAATCTGACAAAGAGCAAGCTTAAAGTCCAGAGGTCGCC 1915  
 |||  
 Db 173 TCAGAAATATGAACAAATCTGACAAAGAGCAAGCTTAAAGTCCAGAGGTCGCC 114  
 |||  
 QY 1916 TGGGCTGATTTGAATATACCTTCCCTGTCATTTTCCATCAGATAGAGACTTGAC 1975  
 |||  
 Db 113 TGGGCTGATTTGAATATACCTTCCCTGTCATTTTCCATCAGATAGAGACTTGAC 54  
 |||  
 QY 1976 CTGTAAAGCTGCCATCTTCTTAACTAAATTTCCCAATAAGAAAAA 2028  
 |||  
 Db 53 CTGTAAAGCTGCCATCTTCTTAACTAAATTTCCCAATAAGAAAAA 1

RESULT 4  
 AL709927 668 bp mRNA linear EST 22-MAR-2002  
 LOCUS AL709927  
 DEFINITION DKFP68602364.f1 666 (synonym: hlcc3) Homo sapiens cDNA clone  
 AL709927  
 VERSION AL709927  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann  
 S.  
 EST (Duesterhoeft, et al.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Duesterhoeft A  
 MFS  
 Am Klopferstr. 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert.  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing  
 consortium of the German Genome Project.  
 No 5' sequence available.  
 This clone (DKFP68602364) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

source  
 1..668  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="DKFP68602364"  
 /clone\_1ib="686 (synonym: hlcc3)"  
 /tissue\_type="human skeletal muscle"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pTRi-plex2; Site\_1: SfiI; Site\_2: SfiI;  
 cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others  
 ORIGIN  
 Query Match 28.6%; Score 581.6; DB 9; Length 668;  
 Best Local Similarity 98.2%; Pred. No. 9,9e-114;  
 Matches 596; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 148 AAGCCGCTTTCGAAGTCTTCCACTTCAGACAGATGCTTACGCTTAAAAAAATGAGCT 207  
 |||  
 Db 60 AAGCCGCAAGCCAGAGTCTTCCACTTCAGACAGATGCTTACGCTTAAAAAAATGAGCT 119  
 |||  
 QY 208 CCGCTGTTTCTGTTTCCAGATGAGCCATCTGGCTGATTTCTTCCATGATGAGCA 267  
 |||  
 Db 120 CCGCTGTTTCTGTTTCCAGATGAGCCATCTGGCTGATTTCTTCCATGATGAGCA 179  
 |||  
 QY 268 CAACATCAGCTCCCTGCTATATGAGGACAGAGCCAGGAGCATGACAGTCTGTTCTGTC 327  
 |||  
 Db 180 CAACATCAGCTCCCTGCTATATGAGGACAGAGCCAGGAGCATGACAGTCTGTTCTGTC 239  
 |||  
 QY 328 TTCTGGGCGCTGGCCGCTTCTTTTGGGGGAGCTTTTGGGACACCGACATGTTT 387  
 |||  
 Db 240 TTCTGGGCGCTGGCCGCTTCTTTTGGGGGAGCTTTTGGGACACCGACATGTTT 299  
 |||  
 QY 388 CTACCTGATGAGCCCGGCTGACGCTGTGATGACCTTCAAGCAGACCGGCTGAT 447  
 |||  
 Db 300 CTACCTGATGAGCCCGGCTGACGCTGTGATGACCTTCAAGCAGACCGGCTGAT 359  
 |||  
 QY 448 GCTGCACATGAGCTGTGCGGAGTGTATAGGGGCGCTTCTTGTGACATGAGGCTTT 507  
 |||  
 Db 360 GCTGCACATGAGCTGTGCGGAGTGTATAGGGGCGCTTCTTGTGACATGAGGCTTT 419  
 |||  
 QY 508 TGATGCTACATGAGCACTGCTGCTCCGAGAGACAGTCCAGCTTTCAGTGGAGACAG 567  
 |||  
 Db 420 TGATGCTACATGAGCACTGCTGCTCCGAGAGACAGTCCAGCTTTCAGTGGAGACAG 479  
 |||  
 QY 568 CCGGGCCCTGTGTCTTGTGACACTGCTGTGATCATTCACCAAGATGAATATCCCCG 627  
 |||  
 Db 480 CCGGGCCCTGTGTCTTGTGACACTGCTGTGATCATTCACCAAGATGAATATCCCCG 539  
 |||  
 QY 628 -GGCTCACTGACAGGCTCCGTGAGTGCACAGAGCCCTTGAAGTGGGGAGAGGCTGCC 686  
 |||  
 Db 540 NGGCTCACTGACAGGCTCCGTGAGTGCACAGAGCCCTTGAAGTGGGGAGAGGCTGCC 599  
 |||  
 QY 687 GCTCTTACAGCCAGCGTGTGCTCAGAGAGGTGCGCTTCTTCAACCTGACGCTCTTAC 746  
 |||  
 Db 600 GCTCTTACAGCCAGCGTGTGCTCAGAGAGGTGCGCTTCTTCAACCTGACGCTCTTAC 659  
 |||  
 QY 747 CGCTGCTGA 755  
 |||  
 Db 660 CGCTGCTGA 668

RESULT 5  
 BM129080 571 bp mRNA linear EST 12-MAR-2002  
 LOCUS BM129080  
 DEFINITION l1f17c04.y1 melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
 cDNA clone IMAGE:5676655 5' similar to TR:0915R3 0915R3  
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ; mRNA sequence.  
 BM129080  
 ACCESSION BM129080.1 GI:17123632  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 571)  
 Lemisha, I., Pearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Bistara, A.,  
 Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cardenas,  
 M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, V., Williams, T.,  
 Jackson, Y. and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: l1f17c04.x1

CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138



Db 44 AACCCAGCCAGCAGCTCTTCCACTTCCAGCAGCAATGCTACTGCTTAAATAATGAAGCT 103  
 QY 208 CCTGCTGTTTCTGTTTCCAGATGAGCCATCTTGCTCTATTTCTCCACATGTACAGCA 267  
 Db 104 CCTGCTGTTTCTGTTTCCAGATGAGCCATCTTGCTCTATTTCTCCACATGTACAGCA 163  
 QY 268 CACATCAGTCCCTGCTCTGATGAGCAGAGCCGAGCCGATCAGCTGCTGTTGTC 327  
 Db 164 CACATCAGTCCCTGCTCTGATGAGCAGAGCCGAGCCGATCAGCTGCTGTTGTC 223  
 QY 328 TTCTGCGCTGCTGCTCTTCTTTTGAGGAGAGCTTTTGGAGACACCCAGATGTTT 387  
 Db 224 TTCTGCGCTGCTGCTCTTCTTTTGAGGAGAGCTTTTGGAGACACCCAGATGTTT 283  
 QY 388 CTACTGATGAGAGCCGCTGAGCAGCTGTGATGACCTTCAACAGAGCAGCCCTGGAT 447  
 Db 284 CTACTGATGAGAGCCGCTGAGCAGCTGTGATGACCTTCAACAGAGCAGCCCTGGAT 343  
 QY 448 GCTGACATGAGCTGTGAGGAGATGATACAGGCGCTCTTCTTGACATGAGCTTT 507  
 Db 344 GCTGACATGAGCTGTGAGGAGATGATACAGGCGCTCTTCTTGACATGAGCTTT 403  
 QY 508 TGATGCTACATGAGACCTGTGCTCCCGAGACAGTCCAGCTCTTTGAGTGGAGAGAG 567  
 Db 404 TGATGCTACATGAGACCTGTGCTCCCGAGACAGTCCAGCTCTTTGAGTGGAGAGAG 463  
 QY 568 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627  
 Db 464 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 523  
 QY 628 GGTCTACATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667  
 Db 524 GGTCTACATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 583  
 QY 688 CTCTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747  
 Db 584 CTCTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643  
 QY 748 GCTCTGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806  
 Db 644 GCTCTGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703  
 QY 807 CCGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 864  
 Db 704 CCGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763  
 QY 865 GATGGGAGAGATGAGAGAGAA--CTCAGAGAGAGAGAG--AACCTACTATGTATG-C 920  
 Db 764 GATGGGAGAGATGAGAGAGAA--CTCAGAGAGAGAGAG--AACCTACTATGTATG-C 823  
 QY 921 AGTTCATCTGCCAAA--GCCAGCTGAGATGTAC--AAGACCATCACTCTTGCCCAAG 977  
 Db 824 AGTTCATCTGCCAAA--GCCAGCTGAGATGTAC--AAGACCATCACTCTTGCCCAAG 883  
 QY 978 CCGTGCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035  
 Db 884 CCGTGCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943  
 QY 1036 CCAGACTTCCC 1046  
 Db 944 CCAGACTTCCC 954

RESULT 3  
 LOCUS BM969292/2 593 bp mRNA linear EST 21-MAR-2002  
 DEFINITION UI-CF-ENO-acp-1-21-0-UI-s1 UI-CF-ENO Homo sapiens cDNA clone  
 ACCSSION BM969292  
 VERSION BM969292.1 GI:19586879  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 593) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996) 9704447 Contact: McCray, PB McCray Lab University of Iowa 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA Tel: 319 356 4866 Fax: 319 356 7171 Email: paul-mccray@uiowa.edu Tissue Procurement: Dr. M. J. Welsh, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 FORWARD PolA=yes Location/Qualifiers 1..593 /organism="Homo sapiens" /db\_xref="taxon:9606" /clone="UI-CF-ENO-acp-1-21-0-UI" /clone\_1b="UI-CF-ENO" /tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells" /dev\_stage="Adult" /lab\_host="DH10B (Life Technologies) (T1 phage resistant)" /note="Organ: Lung; Vector: pRT3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT. TAG\_LIB=UI-CF-ENO TAG\_TISSUE=Human Lung Epithelial Cell Lines untreated LPS 6hr to LPS 24h TAG\_SEQ=CTGCTCAGGT"

BASE COUNT 157 a 127 c 137 g 172 t  
 ORIGIN  
 Query Match 29.1%; Score 591.4; DB 14; Length 593;  
 Best Local Similarity 99.8%; Pred. No. 8.3e-116;  
 Matches 592; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1436 ACATGCTGTGGGATACACACTGAGTGTGATGCTGCTACACAGCTCTAAGCAGAGGA 1495  
 Db 593 ACATGCTGTGGGATACACACTGAGTGTGATGCTGCTACACAGCTCTAAGCAGAGGA 534  
 QY 1496 CTTTGTGCTCATGCTGTGCTAGAAACAGACTGGGAGACCTTATGTAGACAGCAT 1555  
 Db 533 CTTTGTGCTCATGCTGTGCTAGAAACAGACTGGGAGACCTTATGTAGACAGCAT 474  
 QY 1556 CCACACAGTGAACAGAGGATTTCTCTCTTTCTTGATCTTCTGCTGCTGAGAGC 1615  
 Db 473 CCACACAGTGAACAGAGGATTTCTCTCTTTCTTGATCTTCTGCTGCTGAGAGC 414  
 QY 1616 TTGAGAGACTTTGGGCTGAGAGCCATTAAAGCAGACAGACTATCAGAGGATGATC 1675  
 Db 413 TTGAGAGACTTTGGGCTGAGAGCCATTAAAGCAGACAGACTATCAGAGGATGATC 354





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:55:29 ; Search time 2886 Seconds

(without alignments)  
11403.059 Million cell updates/sec

Title: us-09-816-825-1

Perfect score: 1 ggcctcagagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Sequence: 1 ggcctcagagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY\_NUC  
Gap10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estda.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estinu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estcom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_iny.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vtl.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rdd.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	769.6	37.9	1923	11	AK009113 Mus muscu
2	761.4	37.5	1924	13	B1823850 B1823850 603039012
3	551.4	29.1	593	14	BM969292 UI-CF-ENO
4	581.6	28.6	668	9	AL709927 DKFZP6860
5	569.4	28.0	571	13	BM129080 i17c04.Y
6	544.2	26.8	680	12	BF878439 MR0-ET010

C	7	478	23.5	553	13	BM128370	BM128370 i17c04.x
C	8	478	23.5	553	13	BM128831	BM128831 i17c04.x
C	9	384.2	18.9	783	13	BS963298	BS963298 602827716
C	10	368.6	18.1	695	13	BS964671	BS964671 602831875
C	11	364.4	17.9	852	13	BS966340	BS966340 602832825
C	12	361.8	17.8	965	12	BS579746	BS579746 602905056
C	13	332.4	16.4	419	10	BM572390	BM572390 X09A06.X
C	14	324	15.9	902	14	BS947021	BS947021 AGENCOURT
C	15	288	14.2	362	10	AM002418	AM002418 w161903.x
C	16	282.6	13.9	834	14	BS084877	BS084877 AGENCOURT
C	17	275.2	13.5	358	12	BF056640	BF056640 7X10d12.x
C	18	274.4	13.5	695	12	BS857538	BS857538 7G01A08.X
C	19	263.6	13.0	735	12	BS856652	BS856652 7G01A09.X
C	20	260.2	12.8	536	13	BM245312	BM245312 K0722507-
C	21	252.4	12.4	620	9	A1824100	A1824100 w16c01.x
C	22	235.4	11.6	1067	17	CNS03KN7	AL248380 Tetradon
C	23	230.8	11.4	497	13	BM246681	BM246681 K0741E04-
C	24	221	10.9	525	12	BF197521	BF197521 7G84A08.X
C	25	192.6	9.5	849	17	CNS04QFN	AL302540 Tetradon
C	26	182.4	9.0	417	10	AMS72510	AMS72510 xq18011.x
C	27	167	8.2	568	9	A115260	A115260 u14c07.Y
C	28	158.8	7.8	657	17	AG035205	AG035205 Pan trogl
C	29	156.2	7.7	517	12	BF042384	BF042384 BP250022A
C	30	156.2	7.7	656	9	AA522184	AA522184 v198b09.r
C	31	155.2	7.6	433	17	A2405100	A2405100 I00173108
C	32	147.4	7.3	436	10	AM081348	AM081348 xc41b06.X
C	33	143.4	7.1	429	10	BB849113	BB849113 BB849113
C	34	142.2	7.0	295	9	A1614882	A1614882 v198b09.Y
C	35	141.8	7.0	380	13	BM433243	BM433243 I0m10F03
C	36	126.4	6.2	599	13	BJ031352	BJ031352 BJ031352
C	37	123.8	6.1	2778	11	BC017499	BC017499 Homo sap1
C	38	120.6	5.9	961	14	BO734733	BO734733 AGENCOURT
C	39	117.2	5.8	640	9	AL681288	AL681288 AL681288
C	40	114.2	5.6	634	9	AU803328	AU803328 AU803328
C	41	106.6	5.2	1169	14	BO067314	BO067314 AGENCOURT
C	42	92.6	4.6	514	14	BO560553	BO560553 H4065A07-
C	43	92.2	4.5	662	12	BF344303	BF344303 602017370
C	44	88	4.3	2070	11	AK011202	AK011202 Mus muscu
C	45	86.6	4.3	322	9	A1156825	A1156825 u14c08.Y

## ALIGNMENTS

RESULT 1	AK009113	1923 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK009113				
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003618:carboxylate (chondroitin 6/keratan) sulfotransferase 4, full insert sequence.				
ACCESSION	AK009113				
VERSION	AK009113.1	GI:12843701			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone:2310003618.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)				
AUTHORS	JOURNAL	2049374			
MEDLINE	2049374				
PUBMED	11042159				



1256 TACCGCCAGTCAGATCTGAAACAAGAACAGGAAACCTGTTGCTGGATCTTCTG 1309

cc bronchial asthma, hypersensitivity, pneumatic fever and tissue rejection during transplantation.



***This Page Blank (uspto)***



```

QY 1082 CATCTCAGACCTGGTGTGATCAATCATCACCAGGCAAGGCGATGGTG-----ACCAAC 1135
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1584 CAGTGTGAGGCTGTGATCCATACATCAACCCAGGATCTGTGACCTGTGCGCCGCCGGA 1643
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1136 GCTTCCACACAAATGCCAGGAGTCCCTTAATGTCTCCACAGCTTGCGCCTGTCTTTG 1195
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1644 GCCTTCAAGACTTGCTCCAGAAATGCCCTCAAGCTCTCCAGGCTTGCCCAATGCGCTG 1703
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1196 CCCATGAAAGGTTTCTGAGCTTCAGAAAGCCGTGCGCATGCGCATGAATTTGCTGGGC 1255
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1704 CCGTTTGCAAGATCCCGCGGCTGCAAGACTGTGCGCTGTGCGCTGCACTGTGCGGC 1763
    || || || || || || || || || || || || || || || || || || || || || ||
QY 1256 TACCGCCAGCTGAGATCTGACACAGAAACAGAGAAACCTGTGCTGATCTCTG 1309
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1764 TACCGGCGCTGTGACTCTGAGAGACGACGAGGCAACCTGCGCCCTGTGATCTGTG 1817
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 13
AAD02697
ID AAD02697 standard; DNA: 160552 BP.
XX
AC AAD02697;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
XX
KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ds.
XX
OS Homo sapiens.
XX
FH Key
FH exon
FT 32847..32922
FT /tag= a
FT /number= 1
FT /label= 4a.504
FT 32923..35592
FT /tag= b
FT /cons_splice= (5'site:NO, 3'site:YES)
FT 35593..35674
FT /tag= c
FT /number= 2
FT /label= 4a.503
FT 35675..45093
FT /tag= d
FT 45094..45185
FT /tag= e
FT /number= 3
FT /label= 4a.502
FT 45186..46633
FT /tag= f
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 46634..46700
FT /tag= g
FT /number= 4
FT /label= 4a.501
FT 46701..47938
FT /tag= h
FT /cons_splice= (5'site:YES, 3'site:NO)
FT 47939..49746
FT /tag= i
FT /number= 5
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"

```

```

FT 5'UTR
FT 47939..47955
FT /tag= j
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 47956..49128
FT /tag= k
FT /product= "Human glycosyl transferase-4alpha
FT (GST-4alpha)"
FT 49129..49746
FT /tag= l
FT 83257..83347
FT /tag= m
FT /label= 4a.502
FT 83348..96412
FT /tag= n
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 96413..96484
FT /tag= o
FT /label= 4a.501
FT 96485..98456
FT /tag= p
FT /cons_splice= (5'site:NO, 3'site:NO)
FT 98457..99968
FT /tag= q
FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT 98457..98473
FT /tag= r
FT /note= "Portion of 5' untranslated region (5'UTR)"
FT 98474..99661
FT /tag= s
FT /product= "Human glycosyl transferase-4beta
FT (GST-4beta)"
FT 99662..99968
FT /tag= t
FT
XX PN MO200106015-A1.
XX PD
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000NC-US19741.
XX
XX 20-JUL-1999; 990S-0144694.
XX 13-JUL-2000; 2000DS-0593828.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S.
XX WPI: 2001-138471/14.
XX P-PDB; AAY72639, AAY72640.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications -
XX
XX Example 1; Page 62-104; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
XX DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
XX chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX CC disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

```

```

DB 780 GTGACACTGCTGCGGACCGCGGGCGCGCGCTCCCGGAGACAGACACCGCAAGCT 839
QY 842 CTCATGATTGACAGTCGATCTGTGATGGGCGACATGAGCAGACAACTCAAGAGAGAC 901
DB 840 CTGGGCGGTGACAGCGCATCTGCTGGGACCAACGCGACCTGGGTGAGGCGCACCC 899
QY 902 CAACCCCTATGATGTGATGAGTCATGTCGCAAGCCAGCGTGCATGATCTACAGACATC 961
DB 900 GGCCCTGGCGGTGCTGGCGAGGTGTGCCGTAGCCACGTACGCATGCGGAGGCCGACCA 959
QY 962 CAGTCCTTGGCCCAAGGCCCTCGAGAACGCTACCTGCTTGTGGCGCTATGAGAGACTGCT 1021
DB 960 CTCAGAGCCGCGACCTTTCTGCGCGCGCTACCCCTGTGGTCCCTTCAGAGACTTGGCG 1019
QY 1022 CGAGCCCTGTGGCGCGACCTTCCGGAATGATGATGATGATGATGATGATGATGATGAT 1081
DB 1020 CGGAGAGCCGCTGCGAGAAATCCGTCGCTCTACGCTTCTACCTGAGCTACGCTCA 1079
QY 1082 CATCTTCAGACCTGGTGCATATACATCACCCAGCGACGCGCATGAGGCTG-----ACCAC 1135
DB 1080 CAGCTCGAGGCGCTGATGCATACATCACCCAGCGATCTGAGCCGTGGCGCGCGGCA 1139
QY 1136 GCTTTCACACAAATGCCAGGATGCTTAACTGCTCCAGGCTTGGCGCTGCTGCTTG 1195
DB 1140 GCTTTCAGACTTCTGCTCAGAGATGCGCTCAACGCTCCAGGCGCTGGCGCATGCGCTG 1199
QY 1196 CCGTATGAAAGTTCTGACTTCAGAAAGCTGTGGCGATGCGCATGATTTGGCTGGGC 1255
DB 1200 CCGTTTCCGAAAGATCCGCGCGGTGAGGAGACTGTGCGCTGCTGCTGCTGCTGCTG 1259
QY 1256 TACGCGCCAGTCAGATCTGAACAGACAGAGAAACCTGTTGCTGATCTTCTG 1309
DB 1260 TACCGGCTGTGTACTCTGAGGAGAGCAGCGCAACCTGCGCTGTGATCTGCTG 1313

RESULT 12
ABN89506
ID ABN89506 standard; cDNA: 2544 BP.
XX
AC ABN89506;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human corneal N-acetylglyucosamine-6-sulfotransferase cDNA SEQ ID NO:1.
XX
XX Human; N-acetylglyucosamine-6-sulfotransferase; enzyme; GlnNA6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 693..1880
FT /tag="a
FT /product="N-acetylglyucosamine-6-sulfotransferase"
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-0927602.
XX
XX 11-AUG-2000; 2000US-325773P.
XX
XX (FUKU/) FUKUDA M N.
XX PA (AKAM/) AKAMA T O.
XX PI Fukuda MN, Akama TO;
XX
XX WPI: 2002-507643/54.
XX DR P-PSDB: ABB81554.
XX

```

```

PT New nucleic acid encoding corneal
PT N-acetylglyucosamine-6-sulfotransferase, useful for treatment,
PT monitoring and diagnosis of macular corneal dystrophy
XX
XX Claim 4; Fig 1A-D; 69pp; English.
XX
CC The present sequence encodes human corneal
CC N-acetylglyucosamine-6-sulfotransferase (1), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (1) is located to chromosome 16q22,
CC and has ophthalmological activity. (1) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (1) makes possible treatment
CC of MCD without requiring keratolimphasty or keratectomy.
XX
SQ Sequence 2544 BP; 460 A; 800 C; 733 G; 551 T; 0 other:
Query Match 20.4%; Score 414.4; DB 24; Length 2544;
Best Local Similarity 64.5%; Pred. No. 6e-110;
Matches 654; Conservative 0; Mismatches 351; Indels 9; Gaps 2;
QY 302 GAGGCGATGACGCGTGTCTGCTCTCTGCGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 361
DB 807 GCGGCGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 866
QY 362 CTTTGTGGGAGACCCAGATGTTTCTTACCTGATGAGCGCGCTGGCAGCGTGGATG 421
DB 867 CTCTTCACACAGACACCCCGAGCTCTTCTACCTATGAGCGCGCGTGGACGCTGTGAC 926
QY 422 ACCCTCAGCAGACACCGCTGATGCTGACATGCTGTGGGATCTGATATGCGGCC 481
DB 927 ACCCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 986
QY 482 GCTCTCTGTGGACATGACGCTTGTGATGCTGATGATGATGATGATGATGATGATGATGAT 541
DB 987 GTCTTCTGTGGACATGACGCTTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1043
QY 542 TCCAGCTCTTTCAGTGGAGAGAACAGCCGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTG 601
DB 1044 TCCGACCTTTCAGTGGAGAGAACAGCCGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTG 1103
QY 602 ATCCCAACATATATATATATATATATATATATATATATATATATATATATATATATAT 661
DB 1104 TTTCCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1163
QY 662 TTTGAGTGTGAGAGAGCGCTGCTGCTTACAGCCAGCTGCTGCTGCTGCTGCTGCTGCTG 721
DB 1164 TTTGAGTGTGAGAGAGCGCTGCTGCTTACAGCCAGCTGCTGCTGCTGCTGCTGCTGCTG 1223
QY 722 TTTTCAACCTGAGTCTGCTTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
DB 1224 TTTTCAACCTGAGTCTGCTTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1283
QY 782 GTGACCTGTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841
DB 1284 GTGACCTGTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1343
QY 842 CTCATGATTGACAGTCGATCTGTGATGGGCGACATGAGCAGAAATCTCAAGAGAGAGAC 901
DB 1344 CTGGGCGGTGACAGCGCATCTGCTGGGACCAACGCGACCTGGGTGAGGCGCACCC 1403
QY 902 CAACCCCTATGATGTGATGAGTCATGTCGCAAGCCAGCGTGCATGATCTACAGACATC 961
DB 1404 GGCCCTGGCGGTGCTGGCGAGGTGTGCCGTAGCCACGTACGCATGCGGAGGCCGACCA 1463
QY 962 CAGTCCTTGGCCCAAGGCCCTCGAGAACGCTACCTGCTTGTGGCGCTATGAGAGACTGCT 1021
DB 1464 CTCAGAGCCGCGACCTTTCTGCGCGCGCTACCGCGCTGTGCTGCTGCTGCTGCTGCTG 1523
QY 1022 CGAGCCCTGTGAGCCCGACCTTCCGGAATGATGATGATGATGATGATGATGATGATGAT 1081
DB 1524 CGGAGCCGCTGCGAGAAATCCGTGCGCTCTACGCTTCACTGCGCTGATGCTCAGCGCA 1583

```





```

QY 422 ACCTTCAGCAGACACCCGCGCTGATGCTGCACATGGCGCTGCGGATCTGCATACGCGCC 481
DB 1396 ACCCTGTGCGAGGCGACCGCGGACACGCTGCACATGGCGCTGCGGATCTGCATACGCGCC 1455
QY 482 GTCCTCTTGTGCGACATGACGCGCTTGTGATGCTACATGAGAACCTGGTCCCGAGACAG 541
DB 1456 ATCTTTTGTGCGACATGAGACGCTTGTGATGCTACATGAGAACCTGGTCCCGAGACAG 1515
QY 542 TCCAGCCTCTTGTGATGCGGAGACAGCGCGCGCTGTGTCTGCACTGCGCTGTGACATC 601
DB 1516 TCCAGCCTCTTGTGATGCGGAGACAGCGCGCGCTGTGTCTGCACTGCGCTGTGACATC 1575
QY 602 ATCCCAAGATGAATCAATCCCGCGGCTCAGTCAAGGCTCCCTGCTCAATCAACAGCC 661
DB 1576 ATCCCAAGATGAATCAATCCCGCGGCTCAGTCAAGGCTCCCTGCTCAATCAACAGCC 1635
QY 662 TTGAGGTGTGAGAAAGCCCTGCGCTCTACAGCCAGCTGTGCTCAAGAGGTGCGC 721
DB 1636 TTGAGGTGTGAGAAAGCCCTGCGCTCTACAGCCAGCTGTGCTCAAGAGGTGCGC 1695
QY 722 TTCTTCAACCTGAGTCCCTCTACCGCTGTGAAAGACCCCTCCCTCAACCTGCAATATC 781
DB 1696 TTCTTCAACCTGAGTCCCTCTACCGCTGTGAAAGACCCCTCCCTCAACCTGCAATATC 1755
QY 782 GTGCACTGTGTCGCGGACCCCGGCGGTGTTCGCTCCGAGAAACGACAAAGGAGAT 841
DB 1756 GTGCACTGTGTCGCGGACCCCGGCGGTGTTCGCTCCGAGAAACGACAAAGGAGAT 1815
QY 842 CTCATGATGACATGCTGCTGTATGCGGAGCATATGACAGAAATCAAGAGAGAGAC 901
DB 1816 CTCGACAGCGCAACGGATGTCTGTGCGCAACACGCGCAATGGGAGGCGGACCT 1875
QY 902 CACCCCTACTGATGATGACAGTCTATGCGCAAGCCAGCTGAGATCTACAGAACCATC 961
DB 1876 CACCTGCGCGCTGATGCGGAGGTGTGCGGACGACGTCGCGCAATGGGAGGCGGACCA 1935
QY 962 CAGTCTTGTCCCAAGCCCTGCAAGAGACGCTACTGCTGTGCGCTGTGAGAGACCTGCT 1021
DB 1936 CTCAAGCGCGCCCTCTCTCTGCGGCGGCGCTGCGCTGCGCTGCGGAGACCTGCGG 1995
QY 1022 CGAGCCCTTGTGCGGCAAGCTTCCGATGATGATTCGCGGATGATTCGCTTCC 1081
DB 1996 CGGAGCGCGCTGCGGAGATCCGCGGCACTTACCCCTTACCGGCGCTGACCCCTACGCCA 2055
QY 1082 CATCTTCAAGCTGAGGTGCATATACATCACCGGAGCAAGGCGATGGTGA-----CCAC 1135
DB 2056 CAGCTCGAGGCTGATCACACATCACCCGCGGCTGCGGATGCGGAGCAATGAG 2115
QY 1136 GCTTCCACACAATGCGAGGATGCGCTTATGTCGCCAGGCTTGGCGCTGCTTCTTG 1195
DB 2116 GCTTCCACACTTGTGCTGCAAAATGCGGCAAGCTTCCAGGCTTGGCGCTGCGGAGCGCTG 2175
QY 1196 CCGATGAAAGGTTTCTGCACTTCAAGAAAGCTGTGCGATGCCATGAATTCGAGGC 1255
DB 2176 CCGTCACTAAGATCTCGCGGTGAGAGAGGTGTGCGCGGCGCGCTGCACTGCTGGGC 2235
QY 1256 TACCGCCAGCTGATCTGAACAAAGAAAGAAACCTGTGCTGAGATCTCTG 1309
DB 2236 TACCGCGCTGTGTACTGTGCGAGACGAGCGTACTTACCCCTGAGATGTGTG 2289

```

RESULT 9  
ID ABR54724 standard; cDNA: 517 BP.

XX ABR54724;  
AC ABR54724;  
DT 18-JUN-2002 (first entry)  
XX Human colon cancer-associated cDNA, SEQ ID No 194.  
XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

```

XX OS Homo sapiens.
XX PN WO200212280-A2.
XX PD 14-FEB-2002.
XX PE 30-JUL-2001; 2001WO-US23826.
XX PR 03-AUG-2000; 2000US-223265P.
XX PR 02-OCT-2000; 2000US-237406P.
XX PR 20-MAR-2001; 2001US-277495P.
XX PR 03-JUL-2001; 2001US-302702P.
XX PA (CORI-) CORIXA CORP.
XX PI Pyle RA, Xu J, Secretist H;
XX DR WPI: 2002-257462/30.
XX PT Novel polynucleotide encoding colon tumour polypeptides, useful as
XX PS vaccines for treating colon cancers
XX PS Claim 1, page 206; 425pp; English.
XX CC The invention relates to isolated polynucleotides (I) encoding colon
XX CC tumour polypeptides (II). (I) is useful for stimulating an immune
XX CC response in a patient and treating colon cancer in a patient.
XX CC Oligonucleotides derived from (I) are useful for determining the presence
XX CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
XX CC compositions, e.g. vaccines, and other compositions for the diagnosis
XX CC and treatment of colon cancer. A composition comprising a first component
XX CC selected from physiologically acceptable carriers and immunostimulants,
XX CC and an antigen-presenting cell expressing (II) is useful for inhibiting
XX CC development of cancer in a patient. (I) is useful in the design and
XX CC preparation of ribozyme molecules for inhibiting expression of tumour
XX CC polypeptides and (I) ABR54531-ABR55464 represent human colon cancer cDNA
XX CC sequences of the invention.
XX SO Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

Query Match 24.9%; Score 506.4; DB 24; Length 517;
Best Local Similarity 99.8%; Pred. No. 3.3e-137;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1352 GAAGGCTTGTCTCCACACCTGCTGTGAGCTCACTTCTGTGATGCTTGTAGGCT 1411
DB 10 GAGGCGTTTGTCTCCACCTGCTGTGAGCTCACTTCTGTGATGCTTGTAGGCT 69
QY 1412 TGCTTACATCTGAGGCTTAACTACATGCTGTGCTATACACAGAGTGAGTGTG 1471
DB 70 TGCTTACATCTGAGGCTTAACTACATGCTGTGCTATACACAGAGTGAGTGTG 129
QY 1472 TCCACAGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTGTAGAAAAGACTG 1531
DB 130 TCCACAGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTGTAGAAAAGACTG 189
QY 1532 GGGAGCCTTGTGAGAGAGACATCCACAGAGAAAGAGGATGAGCTTCTTTT 1591
DB 190 GGGAGCCTTGTGAGAGAGACATCCACAGAGAAAGAGGATGAGCTTCTTTT 249
QY 1592 CTGATCTTCTGTGCGGAGACTTCAAGAGACTTGTGGCTGAGAGCTATTAAGCAC 1651
DB 250 CTGATCTTCTGTGCGGAGACTTCAAGAGACTTGTGGCTGAGAGCTATTAAGCAC 309
QY 1652 GACACAGTATCAGTGGATGTGATCCATAAACCCTGCTCCATCTTGCCCAATGGGA 1711
DB 310 GACACAGTATCAGTGGATGTGATCCATAAACCCTGCTCCATCTTGCCCAATGGGA 369
QY 1712 ATGATCTTTCACCAAGAGACTCAGCAGATTTTCCACAGAGATGCAATTTGAGCCCT 1771
DB 370 ATGATCTTTCACCAAGAGACTCAGCAGATTTTCCACAGAGATGCAATTTGAGCCCT 429

```

Db 745 CAGGCCCGGAAACAGTCACGCTCTTCCATGGGAGCAAGACCGGCGCTGTGCTCAG 804  
 QY 585 CACCTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTACTGACGCTCC 644  
 Db 805 CCGCTGTGTACTCTTCCCTGCGCCAGGATCAGTCCACCAAGACACGACGACGCTGC 864  
 QY 645 TGTGAGTCAACAGCGCTTTGAGGTGGAGAGGCGCTGCGGCTCCACAGCAGCAGTGG 704  
 Db 865 TGTGGGGTACAGAGCCCTTGTATGATGAGGAGGAGGCGCTGCGCTCAGCGCTTCTG 924  
 QY 705 TGTCTAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTCTACCGCTGTGAAGACCCCT 764  
 Db 925 TACTCAAGAGAGTGGCTTTCTCAAGCCCTGACGCGCTCTATCCACTACGACGCTT 984  
 QY 765 CCGCTAACCTGATATCTGTCACCGCTGCTGCGGAGCCCGGCGCTGTGCTGCTCCGAG 824  
 Db 985 CCGCTAACCTGACGCTGCTGTCACCGCTGCTGCGGAGCCCGGCGCTGTGCTGCTCCGAG 1044  
 QY 825 AACGCACAAGGAGAGATCTCATGATTGACAGTCCGCTTGTATGGGCGACGATGAGAGA 884  
 Db 1045 AGCAACACCATAGAACTCATGTGTGACAGTCAATGATGCTAGGGGAGCATTTGGAAA 1104  
 QY 885 AACGCAAGAGAGAGACCAACCTACTATGTGATGAGAGAGTCACTGCGCAAGCGACGCTGG 944  
 Db 1105 CGATTAAGAGAGAGACCAACGCTTATTTAGCCATGAGAGATATCTGCAAAAGCCAGGTGG 1164  
 QY 945 AGATCTACAGAGACCATCTGCTTGGCCAGGCGCTGACAGAAAGCGTACCTGCTTGTGC 1004  
 Db 1165 ACATAGTCAAGGCGCATCAAAACCGCTCCCTGAGCTGTGACACAGCGCTACCTGTTCTCTGA 1224  
 QY 1005 GCTATGAGAGAGCTGGCTGAGACCCCTGTGGCCAGACTTCCCGAATGATGATTCGTGG 1064  
 Db 1225 GGTATGAGAGAGCTGGCTGGGACCCCTGCGCCAGAGAGACGACATATATTAATTTGTGG 1284  
 QY 1065 GATTGAGATCTTCCCGCATCTTGTGAGAGCTGTGATTAACATACCCGAGGAGCAAGGCA 1124  
 Db 1285 GATTGAGATCTTCCCGCATCTTGTGAGAGCTGTGATTAACATACCCGAGGAGCAAGGCA 1344  
 QY 1125 TGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTATGTCCTCCAGGCTTGGC 1184  
 Db 1345 TGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTATGTCCTCCAGGCTTGGC 1404  
 QY 1185 GCTGCTCTTCCCGCATCTTGTGAGAGCTGTGATTAACATACCCGAGGAGCAAGGCA 1244  
 Db 1405 GCTGCTCTTCCCGCATCTTGTGAGAGCTGTGATTAACATACCCGAGGAGCAAGGCA 1464  
 QY 1245 ATTGCTGGAGTACCGGACGCTCAGATCTGACACAGAGAGAGAACTGCTGTGATGATC 1304  
 Db 1465 ATTGCTGGAGTACCGGACGCTCAGATCTGACACAGAGAGAGAACTGCTGTGATGATC 1524  
 QY 1305 TTTGCTGCTACCTGAGCTGTCCCTGACAAATCCACTAAGAGGTTGAGAGGCTTTCTG 1364  
 Db 1525 TTTGCTGCTACCTGAGCTGTCCCTGACAAATCCACTAAGAGGTTGAGAGGCTTTCTG 1584  
 QY 1365 --CCACCTGTGTGAGGCTCAGTAC 1388  
 Db 1585 CACCCCTGTGTGAGGCTCAGTAC 1610

## RESULT 8

AAC76156  
 ID AAC76156 standard; cDNA; 2988 BP.

AAC76156;

DT 08-FEB-2001 (first entry)

Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.

Human: open reading frame; ORFX: detection; cytosolic; hepatotropic;  
 KM: vulnary; antiparkinsonian; antiparkinsonian; neuroprotective;  
 KM: anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;

KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KM hypotensive; dermatological; immunosuppressive; antineuritic;  
 KM antiviral; antibacterial; antifungal; antineuritic; antineuritic;  
 KM antineuritic; gene therapy; cancer; proliferative disorder; hypertension;  
 KM neurodegenerative disease; osteoarthritis; graft vs host disease;  
 KM cardiovascular disease; diabetes mellitus; hypochloridism; SCID; AIDS;  
 KM cholesterol ester storage; systemic lupus erythematosus; infection;  
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KM bone damage; cartilage damage; antineuritic disease; coagulation;  
 KM thrombosis; contraceptive; ss.  
 OS Homo sapiens.  
 PN WO200058473-A2.  
 PD 05-OCT-2000.  
 PF 31-MAR-2000; 2000WO-US08621.  
 PE 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 PA (CURA-) CURAGEN CORP.  
 PI Shimkels RA, Leach M.  
 DR WPI: 2000-602362/57.  
 DR P-PSDB: AAB41947.  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 5; Page 2597-2599; 5507pp; English.  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3151. The ORFX  
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;  
 CC antiparkinsonian; antiparkinsonian; neuroprotective;  
 CC osteoproliferative; antineuritic; antineuritic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineuritic; antineuritic; antineuritic; antineuritic;  
 CC antineuritic; antineuritic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypochloridism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antineuritic disease, to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;

Query Match 27.0%; Score 548.8; DB 21; Length 2988;

Best Local Similarity 72.1%; Pred. No. 3.8e-149;

Matches 731; Conservative 0; Mismatches 277; Indels 6; Gaps 1;

QY 302 GAGCGCATGACAGTGTGCTGTCTTCTGCGGCTGTGCTTCTTTGAGGAG 361  
 Db 1276 GATCGTGTGACAGTGTGCTGTCTTCTGCGGCTGTGCTTCTTTGAGGAG 1335  
 QY 362 CTTTGTGGGACAGCCCAATGTTTCTACCTGATGAGAGCCCGCTGACAGTGTGATG 421  
 Db 1336 CTTTGTGGGACAGCCCAATGTTTCTACCTGATGAGAGCCCGCTGACAGTGTGATG 1395

XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;  
SO Query Match 39.6%; Score 805.6; DB 22; Length 877;  
Best Local Similarity 98.3%; Pred. No. 2e-224;  
Matches 834; Conservative 0; Mismatches 11; Indels 3; Gaps 2;  
QY 114 ACTTACACACTTCCTGGGAGGAGGCTTTCACAGCCCGCTTTCAGAGGTCTCCACT 173  
DB 1 ACTTTCACACTTCCTGGGAGGAGGCTTTCACAGCCCGCTTTCAGAGGTCTCCACT 60  
QY 174 TCAGCAATGCTACTGCTTAAATAAATGAAGCTGCTGTTCTTCCAGATG 233  
DB 61 TCAGCAATGCTACTGCTTAAATAAATGAAGCTGCTGTTCTTCCAGATG 120  
QY 234 CCATCTTGCTATCTTCTGACATGATGACGACCAACATCAGCTCCGTCTATGAAG 293  
DB 121 CCATCTTGCTATCTTCTGACATGATGACGACCAACATCAGCTCCGTCTATGAAG 180  
QY 294 CACAGCCGAGCCGACATGACGCTGCTGCTTCTTCTGAGCTGCTGCTCTTTTG 353  
DB 181 CACAGCCGAGCCGACATGACGCTGCTGCTTCTTCTGAGCTGCTGCTCTTTTG 240  
QY 354 TGGGGACGCTTTTGGGACACCCAGATGTTTCTAAGTGAGGCCGCTGCGACG 413  
DB 241 TGGGGACGCTTTTGGGACACCCAGATGTTTCTAAGTGAGGCCGCTGCGACG 300  
QY 414 TGTGATGACCTTCAACGACGACCCGCTGATGCTGACATGGCTGTCGGGATCTGA 473  
DB 301 TGTGATGACCTTCAACGACGACCCGCTGATGCTGACATGGCTGTCGGGATCTGA 360  
QY 474 TACGGGCGCTTCTTGTGTGACATGAGCTCTTGTATGCTTACATGAACTGTGCCCC 533  
DB 361 TACGGGCGCTTCTTGTGTGACATGAGCTCTTGTATGCTTACATGAACTGTGCCCC 420  
QY 534 GGGAGACGTCCAGGCTCTTCAAGTGGGAGAACGCGGGCGCTGTTGCACTGCGCT 593  
DB 421 GGGAGACGTCCAGGCTCTTCAAGTGGGAGAACGCGGGCGCTGTTGCACTGCGCT 480  
QY 594 GTGACATCATCCACAGATGAATATATCCCGGGGCTCACTGAGGCTCTGTGAGTC 653  
DB 481 GTGACATCATCCACAGATGAATATATCCCGGGGCTCACTGAGGCTCTGTGAGTC 540  
QY 654 AACGCGCTTGTAGTGTGAGAGGCGCTGCGCTCCTACACGACGTGTGCTCAAG 713  
DB 541 AACGCGCTTGTAGTGTGAGAGGCGCTGCGCTCCTACACGACGTGTGCTCAAG 600  
QY 714 AGTGGCGCTTCTCAACCTCAGTCCCTCTACCGCTGTGAAAGACCCCTCCCTCAAC 773  
DB 601 AGTGGCGCTTCTCAACCTCAGTCCCTCTACCGCTGTGAAAGACCCCTCCCTCAAC 660  
QY 774 TGCATATGCTGACCTGGTCCGGGACCCCGGGCGCTGTTCCTGTCGGAAGACGACA 833  
DB 661 TGCATATGCTGACCTGGTCCGGGACCCCGGGCGCTGTTCCTGTCGGAAGACGACA 720  
QY 834 AGGAGATTCATATGATGACAGTGGCATGAGTGGGACGATGACCAAACTCAGA 893  
DB 721 AGGAGATTCATATGATGACAGTGGCATGAGTGGGACGATGACCAAACTCAGA 780  
QY 894 AGGAGACCAACCTACTATGATGACAGTCA--TCTGCCAAAGCCAGC--TGGAGATCT 950  
DB 781 AGGAGACCAACCTACTATGATGACAGGCAATCTGCCAAAGCCAGCTTGGAGACT 840  
QY 951 ACAAGACC 958  
DB 841 ACAAGAAC 848

RESULT 7  
AAZ20793  
ID AAZ20793 standard: DNA: 1926 BP.  
XX  
AC AAZ20793;

XX 08-DEC-1999 (first entry)  
XX Mouse glycosyl sulfotransferase-3 coding sequence.  
DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;  
KW secondary lymph organ; ss.  
XX Mus sp.  
OS W09949018-A1.  
XX W09949018-A1.  
XX 30-SEP-1999.  
XX 26-FEB-1999; 99WO-US04316.  
XX 20-MAR-1998; 98US-0045284.  
PR 12-NOV-1998; 98US-0190911.  
XX (REGC ) UNIV CALIFORNIA.  
PA (SYNT ) SYNTAX USA INC.  
XX Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;  
PI WPI: 1999-580442/49.  
DR P-PSDB: AAY39919.  
XX Human and murine glycosyl sulfotransferase 3 and related  
PT polynucleotides  
XX Claim 4; Fig 3; 59pp; English.  
PS This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of  
XX the invention. The nucleic acid sequences, probes and primers derived  
CC from these, proteins and antibodies are useful in detecting homologues.  
CC The sequences, antibodies and methods are useful in the diagnosis and  
CC treatment of diseases associated with selectin binding interactions,  
CC including conditions associated with or resulting from the homing of  
CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
XX to secondary lymph organs.  
XX Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;  
SO Query Match 36.5%; Score 741.6; DB 20; Length 1926;  
Best Local Similarity 76.4%; Pred. No. 1.5e-205;  
Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;  
QY 165 TCTTCCACTTTCAGCAATGCTACTGCTTAAATAAATGAGCTCCTGTTCTGTTT 224  
DB 388 TCTTCTTCTTCCGACAGATGATGCTGTTAAGAAAGGAGGCTGCTGATGTTCTGGGTT 447  
QY 225 CCAGATGSCCATTTGCTTATTTCTTCCACATGTACAGCCACAAATCAGTCCCTGT 284  
DB 448 CCAGATGTCATGCTGTAGCTCTCTTATCATATGATGCTCCACAGACACCTTCCAG- 506  
QY 285 CTATGAGGACACAGCCCGAGCGGATGACGCTGCTGTTGTGCTCTCTGCGCTGGCT 344  
DB 507 --AGGAGAGATCCAGAGAGCCCGGATGATGCTGCTGCTGCTCTCTCTCTGCGCTGAGAT 564  
QY 345 CTCTCTTTTGGGCGACGCTTTTGGGACGACCCAGATGTTTCTAAGTGAAGCCCG 404  
DB 565 CTCTCTTTTGGGACAGCTTTTGGGACGACCCGAGATGTTTCTAAGTGAAGCCCG 624  
QY 405 CCTGGACGCTGTGATGATCTTCAACAGACAGACGCGCTGATGCTGACATGAGTGC 464  
DB 625 CCTGGATGATGATGATCTTCAACAGACAGACGCTGAGAGCTGACAGAGGCTGTC 684  
QY 465 GGGATGATAGCGGCGCTCTTCTTGTGACATGAGCGCTTTGATGCTCATGAGAAC 524  
DB 685 GGGATTTTGGCTTCCGCTTCTCTGTGACATGAGCGCTTTGATGCTCATGAGAAC 744  
QY 525 CTGTCTCCCGGAGACAGTCCAGGCTCTTTCAGTGGAGAAACGCGCGGCTGTGTTG 584

```

XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PS
XX Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
SQ

```

```

Query Match      39.68; Score 805.6; DB 22; Length 877;
Best Local Similarity 98.38; Pred. No. 2e-224;
Matches 834; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

```

```

QY 114 ACTTTCACAGCTTCTGAGAGCGAGTCTTCTCAAGCCCGCTTGGCAAGTCTTCACAT 173
Db 1 ACTTTCACAGCTTCTGAGAGCGAGTCTTCTCAAGCCCGCTTGGCAAGTCTTCACAT 60
QY 174 TCAGCACAATGCTACTGCTTCTCAAGTCTTCTCAAGTCTTCTTCTTCTTCTTCTTCT 233
Db 61 TCAGCACAATGCTACTGCTTCTCAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120
QY 234 CCATCTGGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 233
Db 121 CCATCTGGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 180
QY 294 CACAGCCGAGCGCATGCGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 353
Db 181 CACAGCCGAGCGCATGCGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 240
QY 354 TGGGCGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 413
Db 241 TGGGCGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 300
QY 414 TGTGGATGACCTTCAAGCAGAGACCGCGCTGATGCTGACATGCTGTTGGCGGATCTGA 473
Db 301 TGTGGATGACCTTCAAGCAGAGACCGCGCTGATGCTGACATGCTGTTGGCGGATCTGA 360
QY 474 TAGGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 533
Db 361 TAGGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 420
QY 534 GGAAGCAGTCCAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 593
Db 421 GGAAGCAGTCCAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480
QY 594 GTGACATCATCCACAAGATGAATATCCCGGGGCTCACTGAGAGCTCTCTGTCAGTGC 653
Db 481 GTGACATCATCCACAAGATGAATATCCCGGGGCTCACTGAGAGCTCTCTGTCAGTGC 540
QY 654 AAGAGCCCTTTGAGTGTGTGAGAGAGCGCTGCGCTCTTCAACAGCCAGCTGTCCTAAG 713

```

```

Db 541 AACAGCCCTTTGAGTGTGTGAGAGAGCGCTGCGCTCTACAGCCAGCGTGTCTCAAG 600
QY 714 AGGTGGCGCTTCTTCAACCGAGATCCCTTACCCGCTGTGAAGACCCCTCCCTCAAGC 773
Db 601 AGGTGGCGCTTCTTCAACCGAGATCCCTTACCCGCTGTGAAGACCCCTCCCTCAAGC 660
QY 774 TGCATATCTGACACCTGTGTCGGGAGACCCCGGCGGTGTCCGTTCCCGAGAACGACAA 833
Db 661 TGCATATCTGACACCTGTGTCGGGAGACCCCGGCGGTGTCCGTTCCCGAGAACGACAA 720
QY 834 AGGAGATCTCATGATTTGACAGTCCGATTTGTATGAGGCGAGATGACAGAACTCAAGA 893
Db 721 AGGAGATCTCATGATTTGACAGTCCGATTTGTATGAGGCGAGATGACAGAACTCAAGA 780
QY 894 AGGAGACCAACCCCTACTATGATGATGACAGTCA--TCTGCCAAGCCAGC-TGAGATCT 950
Db 781 ANAGAGACCAACCCCTACTATGATGATGACAGGCAATCTGSCCAAGCCAGCTTGAGACT 840
QY 951 ACAAGAC 958
Db 841 ACAAGAAC 848

```

```

RESULT 6
AAK93921
ID AAK93921 standard; cDNA; 877 BP.
XX
XX AAK93921;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human cDNA clone representative sequence, SEQ ID NO: 2381.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
XX Homo sapiens.
XX
XX EP1130094-A2.
XX
XX 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX
XX 11-JAN-2000; 2000JP-0118774.
XX
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PS
XX Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used in the
CC representative sequence from a human clone which was used in
CC the homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

```



XX WPI: 2002-075226/10.  
DR P-PSDB; AAU11274.  
XX  
PT New enzyme, useful for modifying acceptor molecule, comprises an  
PT isolated L-selectin sulfoltransferase-2 that directs expression of  
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or  
PT intestinal GlcNAc 6-sulfoltransferase  
PS  
PS Claim 19, Fig 4, 98pp, English.  
XX  
XX The present invention provides a method of modifying an acceptor molecule  
CC by contacting the acceptor with an isolated  
CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active  
CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The  
CC invention also provides a method of treating or preventing an  
CC L-selectin-mediated condition by reducing the expression or activity of a  
CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done  
CC by administering to the subject an oligosaccharide L-selectin antagonist  
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by  
CC administering antibody material that specifically binds beta1,3gnt,  
CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin  
CC sulfoltransferase-2 (LST-2) also directs MECA-79 antigen expression.  
CC Alternatively, the expression or activity of LST-2 or its active  
CC fragment can be reduced in combination with reducing the expression or  
CC activity of beta1,3gnt. The method is useful for treating L-selectin  
CC mediated conditions such as Crohn's disease and ulcerative colitis,  
CC inflammatory disorders of the skin such as allergic contact dermatitis,  
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type  
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This  
CC sequence represents cDNA encoding human LST-2.  
XX  
SQ Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;  
Query Match 62.1%; Score 1261.6; DB 24; Length 1333;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1264; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 155 TCTTGCAGAGCTTCCACTTACAGCAATGCTACTGCTTAAAAAAGAGCTCTCTG 214  
DB 66 TCCCTAAAGGCTTCCACTTACAGCAATGCTACTGCTTAAAAAAGAGCTCTCTG 125  
QY 215 TTCTGTGTTCCAGATGGCCATCTGGCTATTCTTCCAGATGACAGCAACAATC 274  
DB 126 TTCTGTGTTCCAGATGGCCATCTGGCTATTCTTCCAGATGACAGCAACAATC 185  
QY 275 AGCTCCCTGTATGAAGGACACGCCGAGCATGCACTGCTGTCTCTCTCTG 334  
DB 186 AGCTCCCTGTATGAAGGACACGCCGAGCATGCACTGCTGTCTCTCTCTG 245  
QY 335 CGCTGTGCTCTTCTTTGTGGGAGAGTTTGGGAGACCCAGATGTTTCTACCTG 394  
DB 246 CGCTGTGCTCTTCTTTGTGGGAGAGTTTGGGAGACCCAGATGTTTCTACCTG 305  
QY 395 ATGAGGCCCTGTGACAGCTGTGATACCTTCAAGCAGAGACCGCTGTGATGCTGAC 454  
DB 306 ATGAGGCCCTGTGACAGCTGTGATACCTTCAAGCAGAGACCGCTGTGATGCTGAC 365  
QY 455 ATGGGTGTGGGAGATGATACGGGCGCTCTTGTGGGAGATGAGAGCTTGTGATGCC 514  
DB 366 ATGGGTGTGGGAGATGATACGGGCGCTCTTGTGGGAGATGAGAGCTTGTGATGCC 425  
QY 515 TACATGAGACCTGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACGCCGGGCC 574  
DB 426 TACATGAGACCTGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACGCCGGGCC 485  
QY 575 CTGTGTGTGACACTGCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTCAC 634  
DB 486 CTGTGTGTGACACTGCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTCAC 545  
QY 635 TGCAGGCTCTGTGACAGCAACAGCCCTTTAGAGTGTGAGAGAGCTGCGCTCTCTAC 694  
DB 546 TGCAGGCTCTGTGACAGCAACAGCCCTTTAGAGTGTGAGAGAGCTGCGCTCTCTAC 605

QY 695 AGCCAGCTGTGTGTCAGAGAGGAGGCGCTTCTCAACCTGCATCCCTCTACCCGCTGCG 754  
DB 606 AGCCAGCTGTGTGTCAGAGAGGAGGCGCTTCTTCAACCTGCATCCCTCTACCCGCTGCG 665  
QY 755 AAGAGCCCTCTCCCTCAACCTGCATATGAGACCTGCTGCGGAGACCCCGGCGCTGTC 814  
DB 666 AAGAGCCCTCTCCCTCAACCTGCATATGAGACCTGCTGCGGAGACCCCGGCGCTGTC 725  
QY 815 CGTTCCTCCGAGACGCAACAGGAGATCTCATGATGACAGTGCATGTGTGCGGCGAG 874  
DB 726 CGTTCCTCCGAGACGCAACAGGAGATCTCATGATGACAGTGCATGTGTGCGGCGAG 785  
QY 875 CATGACAGAACTCAAGAGAGAGACCAACCTACTATGTATGATGATGATGATGATGAT 934  
DB 786 CATGACAGAACTCAAGAGAGAGACCAACCTACTATGTATGATGATGATGATGATGATGAT 845  
QY 935 AGCCAGCTGTGATCTACAGACATCCAGTCTTGTCCCAAGAGCCCTGAGAGAGCTTC 994  
DB 846 AGCCAGCTGTGATCTACAGACATCCAGTCTTGTCCCAAGAGCCCTGAGAGAGCTTC 905  
QY 995 CTGCTGTGCGCTATGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCCGAATGAT 1054  
DB 906 CTGCTGTGCGCTATGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCCGAATGAT 965  
QY 1055 GAATTGTGGGATGGAATTTCTTCCCATTTTCAGACCTGGGTGATATACATCACCCGA 1114  
DB 966 GAATTGTGGGATGGAATTTCTTCCCATTTTCAGACCTGGGTGATATACATCACCCGA 1025  
QY 1115 GGCAGAGGATGAGTATGACACGCTTCCAGACAAATGCCAGGATGCCCTTATGTCTCC 1174  
DB 1026 GGCAGAGGATGAGTATGACACGCTTCCAGACAAATGCCAGGATGCCCTTATGTCTCC 1085  
QY 1175 CAGCTTGGGCGCTGCTTTCCTTGAAGAGTTTTCGACTTTCAGAAAGCTGTGCG 1234  
DB 1086 CAGCTTGGGCGCTGCTTTCCTTGAAGAGTTTTCGACTTTCAGAAAGCTGTGCG 1145  
QY 1235 GATGCCATGATTTGCTGCGCTACCGCCAGCTGATGATGACAGAAAGAAACCTG 1294  
DB 1146 GATGCCATGATTTGCTGCGCTACCGCCAGCTGATGATGATGACAGAAAGAAACCTG 1205  
QY 1295 TTGCTGATCTTCTGCTACCTGAGCTGTCCCTGAGCAAAATCCACTAGAGGATGAGAA 1354  
DB 1206 TTGCTGATCTTCTGCTACCTGAGCTGTCCCTGAGCAAAATCCACTAGAGGATGAGAA 1265  
QY 1355 GCGTTTGTGCGCCAGCTGTGATGAGCTGATGATCTTCTGATGATCTTGTGAGCTTGC 1414  
DB 1266 GCGTTTGTGCGCCAGCTGTGATGAGCTGATGATCTTCTGATGATCTTGTGAGCTTGC 1325  
QY 1415 CTACATCT 1422  
DB 1326 CTACATCT 1333  
RESULT 5  
AAK91803  
ID AAK91803 standard; cDNA: 877 BP.  
XX  
XX AAK91803;  
DE 06-NOV-2001 (first entry)  
XX  
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1130094-A2.  
XX  
XX 05-SEP-2001.  
XX  
XX 07-JUL-2000; 2000EP-0114089.  
PF



```
Db 960 TTGCCCAAGACCCCTGGCAGAACGCTACCTGCTGTGGCTATGAGACCTGGCTCGAGCC 1019
QY 1028 CCTGTGGCCAGACTTCCCGATGTATGAATTGGTGGATTGGAATTTTCCCATCTT 1087
Db 1020 CCTGTGGCCAGACTTCCCGATGTATGAATTGGTGGATTGGAATTTTCCCATCTT 1079
QY 1088 CAGACCTGGGTGCTATACATCACCCGAGAGGAGCATGGGTGACACGCTTTCCACACA 1147
Db 1080 CAGACCTGGGTGCTATACATCACCCGAGAGGAGCATGGGTGACACGCTTTCCACACA 1139
QY 1148 AATGCCAGGAGTGGCCCTTAATAGTCTCCAGGCTTGGCGCTGTCTTGTGCCCTATGAAAAG 1207
Db 1140 AATGCCAGGAGTGGCCCTTAATAGTCTCCAGGCTTGGCGCTGTCTTGTGCCCTATGAAAAG 1199
QY 1208 GTTCTGTGACTTCAGAAAAGCTGTGGGATGCCATGAATTTGCTGGGCTACCGCCACGTC 1267
Db 1200 GTTCTGTGACTTCAGAAAAGCTGTGGGATGCCATGAATTTGCTGGGCTACCGCCACGTC 1259
QY 1268 AGATCTGACAGACAGACAGAACCTGTGTGATGCTTGTCTACTGTGACTGTCCCT 1327
Db 1260 AGATCTGACAGACAGACAGAACCTGTGTGATGCTTGTCTACTGTGACTGTCCCT 1319
QY 1328 GAGCAAAATCCACTAGAGGGGTGAGAAAGCTTTGCTGCCACCTGGTGTGACGCTCACTCA 1387
Db 1320 GAGCAAAATCCACTAGAGGGGTGAGAAAGCTTTGCTGCCACCTGGTGTGACGCTCACTCA 1379
QY 1388 CTTTCTGTGATGCTTGTGAGGCTTGGCTGTGATGCTTGTGAGGCTTAACTACTGTCTGTG 1447
Db 1380 CTTTCTGTGATGCTTGTGAGGCTTGGCTGTGATGCTTGTGAGGCTTAACTACTGTCTGTG 1439
QY 1448 GTATCACACTAGTGTAGTGTGTCCACAGCTGCTCAGCAGAGGACTTTTGTGTCCA 1507
Db 1440 GTATCACACTAGTGTAGTGTGTCCACAGCTGCTCAGCAGAGGACTTTTGTGTCCA 1499
QY 1508 TGTGTGTGTCTAGAAAACAGACCTGGGAGACTTATGTGAGCAGACATCCACACAGTGA 1567
Db 1500 TGTGTGTGTCTAGAAAACAGACCTGGGAGACTTATGTGAGCAGACATCCACACAGTGA 1559
QY 1568 ACAGGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1627
Db 1560 ACAGGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1619
QY 1628 GTGGCTGTGAGGCTTATTAACACAGACAGACAGTATGATGATTAATTAATTAATTAAT 1687
Db 1620 GTGGCTGTGAGGCTTATTAACACAGACAGACAGTATGATGATTAATTAATTAATTAAT 1679
QY 1688 TGTCCACATCTTGGCCCATGGGAGATGATCTTCCACCAAGAGCTCACAGCATTTTCC 1747
Db 1680 TGTCCACATCTTGGCCCATGGGAGATGATCTTCCACCAAGAGCTCACAGCATTTTCC 1739
QY 1748 ACAGAGATGCAATTTCTGAGCCCTTGGAGTTCAGAGTGGATTCAGAGAGAGTGGGAA 1807
Db 1740 ACAGAGATGCAATTTCTGAGCCCTTGGAGTTCAGAGTGGATTCAGAGAGAGTGGGAA 1799
QY 1808 CAAGGTGTGAGTCTTACTATGAGCTTGACCATACAGCATATGCGTATAGAAATATGA 1867
Db 1800 CAAGGTGTGAGTCTTACTATGAGCTTGACCATACAGCATATGCGTATAGAAATATGA 1859
QY 1868 AACCAAAATCTGTGACAAAGAGAGAGCTTTAAGTTTCAAGAGTGGCTGGGCTGCATTT 1927
Db 1860 AACCAAAATCTGTGACAAAGAGAGAGCTTTAAGTTTCAAGAGTGGCTGGGCTGCATTT 1919
QY 1928 GAATATGCTCTCCCTGTGCAATTTTCCATGACATAGAAAGCTTTGACCTGTGAAGTGC 1987
Db 1920 GAATATGCTCTCCCTGTGCAATTTTCCATGACATAGAAAGCTTTGACCTGTGAAGTGC 1979
QY 1988 CATGTGTAAATCTAAATTTCCAAATTAAG 2017
Db 1980 CATGTGTAAATCTAAATTTCCAAATTAAG 2009
```

```
ID AAK94229 standard: cDNA; 1979 BP.
XX
AC AAK94229;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2816.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 94UP-0194486.
XX
PR 11-JAN-2000; 2000JP-0118774.
XX
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-52425/58.
DR P-PSDB; AAM93309.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8: SEQ ID NO 2816; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other:
XX
Query Match 93.5%; Score 1900.8; DB 22; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 114 ACTTTCAGACCTTCGCGGAGCAGAGTCTTTCAGAGCCGCTTCAGAGTCTTCCACT 173
Db 1 ACTTTCAGACCTTCGCGGAGCAGAGTCTTTCAGAGCCGCTTTCAGAGTCTTCCACT 60
QY 174 TCAGCAATGCTACTGCTTAAATAAATGAGAGCTCTGCTGTTCTGTGTTTCCAGATGG 233
Db 61 TCAGCAATGCTACTGCTTAAATAAATGAGAGCTCTGCTGTTCTGTGTTTCCAGATGG 120
QY 234 CCATCTTGTGCTATCTTTCACATGATACAGCAACATGAGCTGCTGATGATGAAG 293
Db 121 CCATCTTGTGCTATCTTTCACATGATACAGCAACATGAGCTGCTGATGATGAAG 180
QY 294 CACAGCCGAGCCCATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 353
Db 181 CACAGCCGAGCCCATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 354 TGGGGCAGCTTTTGGGACAGCACCAGATGTTTCTACTGATGAGAGCCCGCTGACAG 413
Db 241 TGGGGCAGCTTTTGGGACAGCACCAGATGTTTCTACTGATGAGAGCCCGCTGACAG 300
```

|||||  
Db 1861 AATATGAAACAAATCTCTGCACAAAGCAAGCTCTTAAGTTCACAGGGTGGCTGGGC 1920  
QY 1921 TGCATTTGAATATCACTTCCCTCTGCATTTTCCCATCATAGAGACTTTGACCTGTG 1980  
Db 1921 TGCATTTGAATATCACTTCCCTCTGCATTTTCCCATCATAGAGACTTTGACCTGTG 1980  
QY 1981 AAGCTGCCATCTGTTAATACTTAAATTTCCCAATATAGAAAAA 2032  
Db 1981 AAGCTGCCATCTGTTAATACTTAAATTTCCCAATATAGAAAAA 2032  
RESULT 2  
AAZ94211  
ID AAZ94211 standard; cDNA; 2065 BP.  
AC AAZ94211;  
XX 19-JUN-2000 (first entry)  
DT 19-JUN-2000 (first entry)  
XX Human transferase TRNSFS-11 cDNA clone 2617407CBL.  
DE  
XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;  
KW inflammation; gastrointestinal disorder; developmental disorder;  
KW genetic disorder; neurological disorder; reproductive disorder;  
KW smooth muscle disorder; immunological disorder; gene therapy;  
KW diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH 174..1334  
FT CDS /\*tag= a  
XX WO200014251-A2.  
XX 16-MAR-2000.  
XX 09-SEP-1999; 99WO-US20989.  
XX 10-SEP-1998; 98US-0150657.  
XX 04-NOV-1998; 98US-0186779.  
XX 11-MAY-1999; 99US-0133642.  
XX (INCYTE PHARM INC.  
PA  
PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;  
PI Hillman JL, Azimzai Y;  
XX WPI; 2000-256996/22.  
XX P-PSDB; AAY79219.  
XX Human transferase proteins useful for preventing, diagnosing and  
PT treating cancers and developmental, gastrointestinal, genetic,  
PT immunological, neurological, reproductive and smooth muscle disorders -  
XX  
XX Claim 9; Page 104-105; 113pp; English.  
XX The present sequence is that of cDNA clone 2617407CBL encoding  
CC human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human  
CC transferase proteins of the invention (see AAY79209-23). The clone  
CC is expressed from gall bladder cDNA library GBLANOT01. TRNSFS-11  
CC is expressed in dermatologic and gastrointestinal tissues,  
CC especially those associated with inflammation and cell  
CC proliferation. It shows homology to mouse N-acetylglucosamine  
CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides  
CC 264-333 or 1272-1331 of the present sequence can be used as a DNA  
CC probe. The new human transferases and polynucleotides can be used  
CC in the diagnosis, prevention and treatment (including gene therapy  
CC and antisense therapy) of cancer, developmental disorders,  
CC gastrointestinal disorders, genetic disorders, immunological  
CC disorders, neurological disorders, reproductive disorders, and  
CC smooth muscle disorders.

XX SQ Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;  
Query Match 97.5%; Score 1981.2; DB 21; Length 2065;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2005; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
QY 9 GCCAGATGCTCCCTCAGTCTGCGGGAATGCTTCTCATTGCTCTCCAGCCACCTC 68  
Db 1 GCCAGATGCTCCCTCAGTCTGCGGGAATGCTTCTCATTGCTCTCCAGCCACCTC 60  
QY 69 AAGCAGTCTCCCTCAGTCTGCGGGAATGCTTCTCATTGCTCTCCAGCCACCTC 128  
Db 61 AAGCAGTCTCCCTCAGTCTGCGGGAATGCTTCTCATTGCTCTCCAGCCACCTC 120  
QY 129 TGGAGCGAGTCTCTCTCAAGCCGCTTGGCAAGCTCTTCCACATTCACACAAATGCTAC 188  
Db 121 TGGAGCGAGTCTCTCTCAAGCCGCTTGGCAAGCTCTTCCACATTCACACAAATGCTAC 180  
QY 189 TGGCTAAAAAATGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 248  
Db 181 TGGCTAAAAAATGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
QY 249 TCTTCCACATGTACACACCAACATCAGCTCCCTGTCTATGAGGACAGCCGAGCGCA 308  
Db 241 TCTTCCACATGTACACACCAACATCAGCTCCCTGTCTATGAGGACAGCCGAGCGCA 300  
QY 309 TGCACGTGCTGCTTCT 368  
Db 301 TGCACGTGCTGCTTCT 360  
QY 369 GGCAGACCCAGATGTTTCTCTACCTATGAGGAGCCGCTGGACAGTGTAGTACCTTCA 428  
Db 361 GGCAGACCCAGATGTTTCTCTACCTATGAGGAGCCGCTGGACAGTGTAGTACCTTCA 420  
QY 429 AGCAGAGCACCGCTGATGCTGACATGCTGTGCGGATGATACGGGCGCTTCT 488  
Db 421 AGCAGAGCACCGCTGATGCTGACATGCTGTGCGGATGATACGGGCGCTTCT 480  
QY 489 TGTGCGACATGAGCGCTTGTGATGCTTACATGAACTGCTCCCGGAGACAGTCCAGCC 548  
Db 481 TGTGCGACATGAGCGCTTGTGATGCTTACATGAACTGCTCCCGGAGACAGTCCAGCC 540  
QY 549 TCTTTCAGTGGGAGAACACCGCGGCTGTCTGTCACCTGCTGTCATCATCATCCAC 608  
Db 541 TCTTTCAGTGGGAGAACACCGCGGCTGTCTGTCACCTGCTGTCATCATCATCCAC 600  
QY 609 AAGATCAATATCCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667  
Db 601 AAGATG-AATCATCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659  
QY 668 GTGGTGAAGAAGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727  
Db 660 TTTTGGAGAAAGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719  
QY 728 AACCTGCAATCTCTTACCCGCTGCTGGAAGACCCCTCCCTCAACCTGATATGCTGAC 787  
Db 720 AACCTGCAATCTCTTACCCGCTGCTGGAAGACCCCTCCCTCAACCTGATATGCTGAC 779  
QY 788 CTGGTCCGGAGACCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847  
Db 780 CTGGTCCGGAGACCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839  
QY 848 ATTGACAGTGCATTTGTATGAGGAGCATGAGCAGAAATCTCAAGAGAGAGACCAACC 907  
Db 840 ATTGACAGTGCATTTGTATGAGGAGCATGAGCAGAAATCTCAAGAGAGAGACCAACC 899  
QY 908 TACTATGTATGCAAGTCTATCTGCCAAGACAGCTGAGATTTAACAAGACATCCAGTCC 967  
Db 900 TACTATGTATGCAAGTCTATCTGCCAAGACAGCTGAGATTTAACAAGACATCCAGTCC 959  
QY 968 TTGCCAAGGCGCTGAGAAAGCTACCTGCTTGGCTATGAGACCTGAGCTGAGACCC 1027

XX Human and murine glycosyl sulfotransferase 3 and related  
PT polynucleotides  
XX  
PS Claim 4; Fig 1: 59pp; English.  
CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of  
CC the invention. The nucleic acid sequences, probes and primers derived  
CC from these, proteins and antibodies are useful in detecting homologues.  
CC The sequences, antibodies and methods are useful in the diagnosis and  
CC treatment of diseases associated with selectin binding interactions,  
CC including conditions associated with or resulting from the homing of  
CC leukocytes to sites of inflammation and the normal homing of lymphocytes  
CC to secondary lymph organs.  
XX  
SQ Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;  
Query Match 100.0%; Score 2032; DB 20; Length 2032;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
0;  
0Y 1 GGCCTGAGGCCAGGAGTCCCTCCAGTCTGGGGGAAATGCTTCTCATTTGCTTCCAG 60  
DB 1 GGCCTGAGGCCAGGAGTCCCTCCAGTCTGGGGGAAATGCTTCTCATTTGCTTCCAG 60  
0Y 61 CCCAGCTCAAGCAGTCTCCCAACCCCTTGAAGTCAAGCAGTCTTAAAGCTTACTTCA 120  
DB 61 CCCAGCTCAAGCAGTCTCCCAACCCCTTGAAGTCTCAAGCAGTCTTAAAGCTTACTTCA 120  
0Y 121 CAGCTTCTGGAGCAGAGTCTTCTCAAGCCGCTTTCAGAGTCTTCCACTTCAGCAC 180  
DB 121 CAGCTTCTGGAGCAGAGTCTTCTCAAGCCGCTTTCAGAGTCTTCCACTTCAGCAC 180  
0Y 181 AATCTCTACTGCTTAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTT 240  
DB 181 AATCTCTACTGCTTAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTT 240  
0Y 241 GGCCTCTATCTTCCACATGTACAGCCACAAACATCAGTCCCTGCTATGAGAGCACGC 300  
DB 241 GGCCTCTATCTTCCACATGTACAGCCACAAACATCAGTCCCTGCTATGAGAGCACGC 300  
0Y 301 CGAGCGATGACGCTGCTGTTCTTCTTCTTCTGCGCTCTTGCTCTTCTTTTGAGGCA 360  
DB 301 CGAGCGATGACGCTGCTGTTCTTCTTCTTCTTCTGCGCTCTTGCTCTTCTTTTGAGGCA 360  
0Y 361 GCTTTTGGGAGCAGCAGATGTTTCTTCACTGATGAGCCGCCCTGGACGATGATACGCC 420  
DB 361 GCTTTTGGGAGCAGCAGATGTTTCTTCACTGATGAGCCGCCCTGGACGATGATACGCC 420  
0Y 421 GACCTTCAAGCAGACACCGCTGATGCTGACATGCTGCTGCGGATGTATACGGGC 480  
DB 421 GACCTTCAAGCAGACACCGCTGATGCTGACATGCTGCTGCGGATGTATACGGGC 480  
0Y 481 CGTCTCTGTGGACATGAGCGCTTGTGATGCTTCACTGATGAGCGCTGCTGCGGATGT 540  
DB 481 CGTCTCTGTGGACATGAGCGCTTGTGATGCTTCACTGATGAGCGCTGCTGCGGATGT 540  
0Y 541 GTCCAGCCTCTTCAAGTGGAGAACACCGCGCCCTGTGTTCTGACACTGCTGTGACAT 600  
DB 541 GTCCAGCCTCTTCAAGTGGAGAACACCGCGCCCTGTGTTCTGACACTGCTGTGACAT 600  
0Y 601 CATCCCAAAATGAATCATCCCGCGGCTCACTGAGGCTCTGTGCAACAGCC 660  
DB 601 CATCCCAAAATGAATCATCCCGCGGCTCACTGAGGCTCTGTGCAACAGCC 660  
0Y 661 CTTGAGGTGGTGAAGAGGCTGCGCTCTTACAGCCAGTGTGCTCAAGAGGTGGC 720  
DB 661 CTTGAGGTGGTGAAGAGGCTGCGCTCTTACAGCCAGTGTGCTCAAGAGGTGGC 720  
0Y 721 CTTCTTCAACTGCAAGTCCCTTACCGCGTGTAAAGACCCCTCCCAACCTGATAT 780  
DB 721 CTTCTTCAACTGCAAGTCCCTTACCGCGTGTAAAGACCCCTCCCAACCTGATAT 780

0Y 781 CGTGCACCTGTGTCCGGGACCCCGCGCTTCCGTTCCGAGACCCCAAGAGGAGA 840  
DB 781 CGTGCACCTGTGTCCGGGACCCCGCGCTTCCGTTCCGAGACCCCAAGAGGAGA 840  
0Y 841 TCTCATGATGACAGTCCGATGTGATGGGAGCAGCATGAGCAAACTCAAGAAAGAGA 900  
DB 841 TCTCATGATGACAGTCCGATGTGATGGGAGCAGCATGAGCAAACTCAAGAAAGAGA 900  
0Y 901 CCAACCTCTATATGTATGATCAGGTATCTGCCAAAGCAGCTGAGATCTACAAACCAT 960  
DB 901 CCAACCTCTATATGTATGATCAGGTATCTGCCAAAGCAGCTGAGATCTACAAACCAT 960  
0Y 961 CCAGTCTTGGCCAGGCGCTCAGAAAGCATGCTGCTGCGCTATGAGACCTGGC 1020  
DB 961 CCAGTCTTGGCCAGGCGCTCAGAAAGCATGCTGCTGCGCTATGAGACCTGGC 1020  
0Y 1021 TCGAGCCCTGTGGCCAGACTTCCGAAATGATGATGATGATGATGATGATGATGATG 1080  
DB 1021 TCGAGCCCTGTGGCCAGACTTCCGAAATGATGATGATGATGATGATGATGATGATG 1080  
0Y 1081 CCATCTTCAAGCTGAGTGCATATATCACCAGAGCAAGGCAATGGTGAACAGCTTT 1140  
DB 1081 CCATCTTCAAGCTGAGTGCATATATCACCAGAGCAAGGCAATGGTGAACAGCTTT 1140  
0Y 1141 CCACAAATGCGAGGAGTGCCTTAATGTCTCCAGAGGCTGCGCTGCTTCCCTTA 1200  
DB 1141 CCACAAATGCGAGGAGTGCCTTAATGTCTCCAGAGGCTGCGCTGCTTCCCTTA 1200  
0Y 1201 TGAAGAGTTTCTGCACTTCAAGAAAGCTGTGGCATGCAATTTGCTGGCTACCG 1260  
DB 1201 TGAAGAGTTTCTGCACTTCAAGAAAGCTGTGGCATGCAATTTGCTGGCTACCG 1260  
0Y 1261 CCACGTCATGTCGACAGACAGACAGAAACCTGTGCTGATCTTCTGCTACCTGGAC 1320  
DB 1261 CCACGTCATGTCGACAGACAGACAGAAACCTGTGCTGATCTTCTGCTACCTGGAC 1320  
0Y 1321 TGTCTGTGACAAATCCACTAAGAGGTTGAGAGGCTTGTGTCACACTGCTGTCAGCC 1380  
DB 1321 TGTCTGTGACAAATCCACTAAGAGGTTGAGAGGCTTGTGTCACACTGCTGTCAGCC 1380  
0Y 1381 TCACTACTTCTCTGAATGCTTCTGAGCTTCCCTACATCTCTGAGCTTAACTACATG 1440  
DB 1381 TCACTACTTCTCTGAATGCTTCTGAGCTTCCCTACATCTCTGAGCTTAACTACATG 1440  
0Y 1441 TCTGTGGATATCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
DB 1441 TCTGTGGATATCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
0Y 1501 GTGTCCATGCTGTGTCTGAAGAAACAGACTGGGAACTTATGTAGACGACATCCAC 1560  
DB 1501 GTGTCCATGCTGTGTCTGAAGAAACAGACTGGGAACTTATGTAGACGACATCCAC 1560  
0Y 1561 CAGTGAACAGGATATGCTCTTCTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1620  
DB 1561 CAGTGAACAGGATATGCTCTTCTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1620  
0Y 1621 AGACTTGTGGCCCTGAGGCTTATTAAGCAGACAGATGATGATGATGATGATGATG 1680  
DB 1621 AGACTTGTGGCCCTGAGGCTTATTAAGCAGACAGATGATGATGATGATGATGATG 1680  
0Y 1681 ACCTGCTGTCACATCTTGGCCCAATGGGAAATGATCTTTTACCAAGAGCTCACAGC 1740  
DB 1681 ACCTGCTGTCACATCTTGGCCCAATGGGAAATGATCTTTTACCAAGAGCTCACAGC 1740  
0Y 1741 ATTTCACAGAGATCAAAATTCAGCCCTTGAAGTCTCCAGTGAATCAAGAGAGAA 1800  
DB 1741 ATTTCACAGAGATCAAAATTCAGCCCTTGAAGTCTCCAGTGAATCAAGAGAGAA 1800  
0Y 1801 GTGGGAACAAGTGTGATGCTTATGAGCTGACATCAAGATGATGATGATGATGATG 1860  
DB 1801 GTGGGAACAAGTGTGATGCTTATGAGCTGACATCAAGATGATGATGATGATGATG 1860  
0Y 1861 AATATGAACAAATCTGTGCAAAAAGCAAGCTCTTAACTTACAGAGTGTGCTGGCC 1920

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:37:37 ; Search time 427 Seconds

(without alignments)  
10716.770 Million cell updates/sec

Title: US-09-816-825-1

Sequence: 1 ggcctgagagccagatgctc.....ataagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: N\_Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2032	100.0	2032	20	AAZ20792 Human glycosyl sul
2	1981.2	97.5	1975	21	AAZ9421 Human transferrase
3	1900.8	93.5	1909	22	AAZ9422 Human full-length
4	1261.6	62.1	1333	24	AAZ16847 Human l-selectin
5	805.6	39.6	877	22	AAZ91803 Human CDNA 5'-end
6	805.6	39.6	877	22	AAZ93921 Human CDNA clone r
7	741.6	36.5	1926	20	AAZ20793 Mouse glycosyl sul
8	548.8	27.0	2988	21	AAZ76156 Human GREY ORF1711
9	506.4	24.9	517	24	AAZ54724 Human colon cancer

10	414.4	20.4	1647	24	AAZ24670 Human drug metabol
11	414.4	20.4	1694	22	AAZ02700 Human glycosyl sul
12	414.4	20.4	2544	24	ABN89506 Human corneal N-ac
13	414.4	20.4	160552	22	AAZ02697 Human glycosyl sul
14	400	19.7	2044	22	AAZ02699 Human glycosyl sul
15	400	19.7	2170	22	AAZ02698 Human glycosyl sul
16	388.8	19.1	1937	22	AAZ16948 Murine intestinal
17	388.8	19.1	1589	22	AAZ02696 Mouse glycosyl sul
18	378.2	18.6	548	22	AAZ92588 Human CDNA 3'-end
19	337.8	16.6	48435	24	ABN89533 Human corneal N-ac
20	287.4	14.1	389	24	AAZ54724 Human colon cancer
21	143.2	7.0	2156	19	AAZ21200 Glycosaminoglycan
22	142.6	7.0	668	24	ABZ021506 Oligonucleotide fo
23	142.6	7.0	668	24	ABZ021507 Oligonucleotide fo
24	133.6	6.6	1458	19	AAZ36418 Keratan sulfate 6
25	130	6.4	2354	18	AAZ45037 Chick chondroitin
26	123.8	6.1	3029	22	AAH17922 Human CDNA sequenc
27	119.4	5.9	2393	24	AAZ61699 Lung small cell ca
28	119.4	5.9	2409	20	AAZ87821 Human N-acetylgluc
29	109.4	5.4	524	24	ABZ021510 Oligonucleotide fo
30	109.4	5.4	524	24	ABZ021511 Oligonucleotide fo
31	106.4	5.2	524	24	ABZ021512 Oligonucleotide fo
32	106.4	5.2	524	24	ABZ021513 Oligonucleotide fo
33	106.4	5.2	2150	20	AAZ87820 Mouse N-acetylgluc
34	106	5.2	668	24	ABZ021508 Oligonucleotide fo
35	106	5.2	668	24	ABZ021509 Oligonucleotide fo
36	89.4	4.4	576	24	ABZ04408 Oligonucleotide fo
37	89.4	4.4	576	24	ABZ04409 Oligonucleotide fo
38	69.6	3.4	576	24	ABZ045406 Oligonucleotide fo
39	69.6	3.4	576	24	ABZ045407 Oligonucleotide fo
40	62.6	3.1	731	20	AAZ24563 Human lung tumor a
41	62.6	3.1	731	21	AAZ65802 Human lung tumor a
42	62.6	3.1	731	24	ABZ18021 Human lung tumour
43	60	3.0	60	24	ABN38074 Human spliced tran
44	59.2	2.9	303	24	ABN38074 Human small cell ca
45	59.2	2.9	303	24	AAZ61848 Lung small cell ca

## ALIGNMENTS

RESULT 1  
AAZ20792  
ID AAZ20792 standard; DNA: 2032 BP.  
XX AAZ20792;  
AC AAZ20792;  
DT 08-DEC-1999 (first entry)  
XX  
XX Human glycosyl sulfotransferase-3 coding sequence.  
DE  
XX  
XX glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;  
KW selectin binding interaction; inflammation; lymphocyte homing; human;  
KW secondary lymph organ; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9949018-A1.  
XX  
XX 30-SEP-1999.  
XX  
XX 26-FEB-1999; 99WO-US04316.  
XX  
XX 20-MAR-1998; 98US-0045284.  
XX 12-NOV-1998; 98US-0190911.  
XX  
XX (REGC) UNIV CALIFORNIA.  
XX (SYNT) SYNTX USA INC.  
XX  
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;  
XX WPI: 1999-580442/49.  
XX P-PsDB: AAZ39918.

```

Db 108326 CGGAGCGCGCTGGCAGAAATCCGTGCGCTCTACGCTTCACTGAGGCTCAGTCTCAGCCCA 108385
QY 1082 CATCTCAGACCTGGTGCATTAACATCACCCGAGGCAAGGCGCATGGGTG-----ACCAAC 1135
Db 108386 CAGCTCGAGGCGCTGGATCCATTAACATCACCCGAGGATCTGACCTGGTGGCCCGCGGAA 108445
QY 1136 GCCTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTGGCGCTGGTCTTTG 1195
Db 108446 GCCTTCAGAGCTTCGTCCAGGAATGCCCTCAAGTCTCCAGGCTGGCGCATGGCGTG 108505
QY 1196 CCCTATCAAAAGCTTCTCGACTTCAGAAAGCCTGTGGCGATGCGATGAATTTGCTGGCG 1255
Db 108506 CCCTTTCAGAGATCCCGCGCGTCCAGGAATGTGGCGCTGGTGGCTGCACTGCTGGCG 108565
QY 1256 TACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTGGTGAATCTTCTG 1309
Db 108566 TACCGCGCTGTGTACTCTGAGAGAGAGAGCAAGCAACCTCGCCTTGATCTGGTG 108619

```

Search completed: January 10, 2003, 21:02:19  
 Job time : 6578 secs





QY 1022 CGAGCCCTGTGSGCCAGACTTCCGANGATGAATTGTGGATTGGAATTTTCCTTCC 1081  
 Db 70368 CGGAGCCGCTGTGCAGAAATCCGTGCGCTCTAGCCCTTCACTTGGGCTCAAGCCCA 70427  
 QY 1082 CATCTTCAGACCTGGGGGATACATACACCCGAGCAAGGAGATGGGTG-----ACCAC 1135  
 Db 70428 CAGCTCGAGGCTTGATTCATACATACACCCGAGATCTGAGACTGTGTCGGCCGCGCA 70487  
 QY 1136 GCTTTCCACAAATGCGAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCTTTG 1195  
 Db 70488 GCCTTCAGACTTCGTCAAGGATGCGCTCAAGCTTCCAGGCTGCGCCATGCGCTG 70547  
 QY 1196 CCTATGAAAGATTTCTGACTTCAGAAAGCTGTGGGATGCCATGATTTGCTGGGC 1255  
 Db 70548 CCTTTGCCAGATCCGCGCGCTGCGAGAACTGTGCTGTGTGCTGCACTGCTGGGC 70607  
 QY 1256 TACCGCCAGCTCAGATCTGAAAGAGAGAGAAACCTGTGTCTGATCTTCTG 1309  
 Db 70608 TACCGGCTGTGTACTGTGAGAGAGAGAGCAACCTGCGCTTGTATCTGTG 70661

RESULT 14  
 AC009163 157337 bp DNA linear PRI 02-NOV-2001  
 LOCUS sapiens chromosome 16 clone Rpl1-77K12, complete sequence.  
 DEFINITION AC009163  
 AC009163 GI:16596526  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 16  
 JOURNAL Unpublished  
 2 (bases 1 to 157337)  
 REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 3 (bases 1 to 157337)  
 REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Nov 2, 2001 this sequence version replaced gi:7689976.

COMMENT  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 Genbank flat file format but are available as part  
 of this entry's ASN.1 file.

-----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 Genbank flat file format but are available as part  
 of this entry's ASN.1 file.

-----  
 FEATURES  
 Source  
 1. 157337  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="Rpl1-77K12"

BASE COUNT 45295 a 36632 c 36308 g 39102 t  
 ORIGIN  
 Query Match 20.4%; Score 414.4; DB 9; Length 157337;  
 Best Local Similarity 64.5%; Pred. No. 8,2e-96;  
 Matches 654; Conservative 0; Mismatches 351; Indels 9; Gaps 2;

QY 302 GAGCGATGACAGTGTGTGTCTCTCTCCGTGGGCTGTGCTCTCTCTTTTGGGGCAG 361  
 Db 93978 GCGCGCTGATGTCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 93919  
 QY 362 CTTTGTGGGAGACACCGATGTTTCTACCTGATGAGAGCCGCTGACAGTGTGATG 421  
 Db 93918 CTCTTCACACAGACCCGAGAGTCTTACCAATATGAGAGCCGAGTGTGAGAGC 93859  
 QY 422 ACCTTCAGACAGACACCGCTGATGCTGACATGCTGTGCGGATCTATACGGGC 481  
 Db 93858 ACCCTGTGCGAGGCGAGGCGCAACGCTGACATGCTGTGCGGATCTGTGCGCTC 93799  
 QY 482 GTCTTCTGTGAGATGAGCGCTTTGATGCTGATGAGAACTGTGCGGAGAGAC 541  
 Db 93798 GTCTTCTGTGAGATGAGCGCTTTGATGCTGATGAGAACTGTGCGGAGAGAC 93742  
 QY 542 TCACAGCTCTTTCAGTGTGAGAGAACAGCGGCGCTGTCTTGTGACACTGTGACATC 601  
 Db 93741 TCCGACCTCTTTCAGTGTGAGAGAACAGCGGCGCTGTCTTGTGACACTGTGACATC 93682  
 QY 602 ATCCACAGATGAAT 661  
 Db 93681 TTTCCTCCAGAGGCGCATGACAGCGAGGCGCTGTGACAGCACTGTGCGCGGACATC 93622  
 QY 662 TTGAGTGTGTGAG 721  
 Db 93621 TTCACCTGTGCGCGGAG 93562  
 QY 722 TTCTTCAACCTGACATCT 781  
 Db 93561 TTCTTCAACCTGACATCT 93502  
 QY 782 GTGACACTGTGTGCGGAG 841  
 Db 93501 GTGACACTGTGTGCGGAG 93442  
 QY 842 GTCATGATGACAGTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 901  
 Db 93441 GTGCGGCGGTGACAG 93382  
 QY 902 CAACCTACTATGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 961  
 Db 93381 GGCCTGCGGT 93322  
 QY 962 CAGCTTCGCCAG 1021  
 Db 93321 CTAACAG 93262  
 QY 1022 CGAGCCCTGTGCGGAG 1081  
 Db 93261 CGGAGCCCTGTGCGGAG 93202  
 QY 1082 CATCTTCAGACCTGGGTGATACATACACCCGAGCAAGGAGATGGTG-----ACCAC 1135  
 Db 93201 CAGCTCGAGGCTTGATTCATACATACACCCGAGATCTGAGACTGTGTCGGCCGCGCA 93142  
 QY 1136 GCTTTCCACAAATGCGAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCTTTG 1195  
 Db 93141 GCCTTCAGACTTCGTCAAGGATGCGCTCAAGCTTCCAGGCTGCGCCATGCGCTG 93082  
 QY 1196 CCTATGAAAGATTTCTGACTTCAGAAAGCTGTGGGATGCCATGATTTGCTGGGC 1255  
 Db 93081 CCTTTGCCAGATCCGCGCGCTGCGAGAACTGTGCTGTGTGCTGCACTGCTGGGC 93022  
 QY 1256 TACCGCCAGCTCAGATCTGAAAGAGAGAGAAACCTGTGTCTGATCTTCTG 1309

```

repeat_region /rpt_family="Alu" /rpt_type=dispersed complement(15292..15589)
repeat_region /rpt_family="Alu" /rpt_type=dispersed 15668..15718
repeat_region /rpt_family="Alu" /rpt_type=dispersed 15977..16098
repeat_region /rpt_family="MR" /rpt_type=dispersed complement(17258..17433)
repeat_region /rpt_family="MR" /rpt_type=dispersed complement(18001..18310)
repeat_region /rpt_type=dispersed complement(18420..18591)
repeat_region /rpt_family="MR" /rpt_type=dispersed 18732..18767
repeat_region /rpt_family="Simple_repeat" /rpt_type=tandem 18967..20202
CDS
/gene="CHST5"
/note="1-GlcNAc6ST; carbohydrate sulfotransferase 5; expressed in the small intestine and colon"
/codon_start=1
/product="N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="AAG25326.1"
/db_xref="GI:11023148"
/translation="MGKRAVPEKVAHSTRPPAARMMLPFRSSRTVTLLLAQTTCLL
LEIRPSPSPAGGDEIRVHVLISSMRSGSSTFGQLFSPHDVFTIMEPMMHWTTLL
SGSASATAHMAVVDLDRSIFLDCMDVPEAVYPOSRLSLAEPFNATSRALCPSPCAKAE
PGTISKSDQVCTICTROPESILARACRSYSHVYLKEVRFNLQVIVLSDPLNLR
IVHVRDRAVYRSREAGPTLARNNGIVLCTNGKMWADPHILIRFVCSYRIRZ
AAATKPPFLKRRYLVREFEDLAREPLAEIRALYAFGLTITLPOLBANINHTIGSGGI
GAFITSPSSFNARNVSQAMNHALPFKILRVGEVCGALQLLGYRPPVSADQDRDL
GLDLVLPSPHPFVSPASPD"
repeat_region /rpt_family="Alu" /rpt_type=dispersed 20504..20806
repeat_region /rpt_family="Alu" /rpt_type=dispersed 20869..20961
repeat_region /rpt_family="L1" /rpt_type=dispersed 20962..21200
repeat_region /rpt_family="MER4_group" /rpt_type=dispersed 21325..21429
repeat_region /rpt_family="MER4_group" /rpt_type=dispersed 21430..21722
repeat_region /rpt_family="Alu" /rpt_type=dispersed 21725..22007
repeat_region /rpt_family="Alu" /rpt_type=dispersed 22008..22040
repeat_region /rpt_family="Simple_repeat" /rpt_type=tandem 22041..22313
repeat_region /rpt_family="MER4_group" /rpt_type=dispersed 22325..22593
repeat_region /rpt_family="Alu" /rpt_type=dispersed 22594..22949
repeat_region /rpt_family="MER4_group" /rpt_type=dispersed 23116..23321
repeat_region /rpt_family="Simple_repeat" /rpt_type=tandem

```

```

repeat_region      23344..23400
                    /rpt_family="MeLR"
                    /rpt_type=dispersed
repeat_region      23433..23456
                    /rpt_family="Simple_repeat"
                    /rpt_type=tandem
                    complement(23457..23720)
                    /rpt_family="Alt"
                    /rpt_type=dispersed
repeat_region      complement(23789..24140)
                    /rpt_family="MER21_group"
                    /rpt_type=dispersed
                    24141..24287
                    /rpt_family="MeLR"
                    /rpt_type=dispersed
                    complement(24308..24420)
                    /rpt_family="MeLR"
                    /rpt_type=dispersed
                    complement(24467..24528)
                    /rpt_family="MER1_type"
repeat_region

```

Query Match	20.4%;	Score 414.4;	DB 9;	Length 71503;
Best Local Similarity	-64.5%;	Pred. No. 7.8e-96;		
Matches 654;	Conservative 0;	Mismatches 351;	Indels 9;	Gaps 2.

QY	302	GAGGCGATGACCTGGTGGTCTGCTTGGTCTTCTTCTGAGGCTGGAGCTCTTCTTTTGGGGGAG	361
Db	69561	GGCGCGTCGATGCTGCTGGTGGTCTGCTCTGCTGCTGAGGCTGGAGCTCTTCTTCTGAGGCGAG	69710
QY	362	CTTTTGGGCGACCCAGCATCTTTTCTACTGATGAGGCCGCTGGACGTGTGGAGT	421
Db	69711	CTCTTCAACGACGACCCGACGCTCTTCTACGTAATGAGAGCCGCGTGGACGCTGGAGCC	69770
QY	422	ACCTTCAGCAGAGCACCGCGCTGATGATGGACATAGAGAGTGTGGGAGTCTGATAGAGGGCC	481
Db	69771	ACCTGTGCGAGGGGCGAGCGCCGCAACGCTGCACATGCTGTGTGGCGACCTGTGTGCGCTCC	69830
QY	482	GTCCTTCTGTGACATGAGCGCTCTTGTATGCTACATGAAACCTGCTCCCGAGACAG	541
Db	69831	GTCCTCTGTGACATGAGCGCTCTTGTATGCTACATG---CTTGGAGCCGCAACCTG	69887
QY	542	TCCAGCCTCTTATGATGGAGAGAACGCGCGCCCTGTGTCTGTGACCTGGCTGTGACATC	601
Db	69888	TCCACCTCTTCTCAGGGGCGCTGAGCGCTGTGACCTGTGTCTGTGACCTGGCTGTGACATC	69947
QY	602	ATCCACAGATGAATATCATCCGCCGCGCTACTGACAGGCTCTGTGACATCAACAGCC	661
Db	69948	TTTCCCGGAGAGGCCCATACAGCAGAGGCGCTGTGACAGGCTCTGTGACATCAACAGCTC	70007
QY	662	TTTGAAGCTGTGAGAGAGGCTGCGGCTCTCTACAGCCAGCTGTGTCTCAAGAGGTGGCC	721
Db	70008	TTTCAACCTGTGCGCGGAGGCGCTGCGCTCTCTACAGCCAGCTGTGTCTCAAGAGGTGGCC	70067
QY	722	TTCTTCAACCTGACATCCCTTACCCGCGCTGAGAAAGCCCTCTCCCTCAACCTGCATATTC	781
Db	70068	TTCTTCAACCTGACATGCTCTTACCCGCTGCTCAGCGACCCCGCTCAACCTGCATATTC	70127
QY	782	GTCACCTGTGTGGGAGCCCGCGCGCTGTTCGTTCTCCGAGAACGACACAAAGGAGAT	841
Db	70128	GTCACCTGTGTGGGAGCCCGCGCGCTGTTCGTTCTCCGAGAGGACGACAGGACGAT	70187
QY	842	CTCATGATTTGACATGTCATGTGATGTGAGGGGAGCATGAGCAGAACTCAGAGAGAGAC	901
Db	70188	CTGACGCTGTACACGCGCATGCTGTGGGACACAAAGGACAGTGTGGGAGGAGCGACCC	70247
QY	902	CAACCCACTATGTATGTGACGTCATCTGCCAAAGCAGGCTGAGATCTCAAGACATC	961
Db	70248	GCGCTGTGGTGTGGGAGGATGTGGCGTGTAGCCAGTACGATACGCGAGGCGCCACA	70307
QY	962	CAGTCTCTGCCAAGGCCCTGAGAGAAAGCTACTGCTGTGTGGCGCTGTAGAGACCTGAGT	1021
Db	70308	CTCAAGCGGCAACCTTTCTGTGGGCGCGCTACCGCTGTGTGGCGCTGTGAGAGCTGAGG	70367

Y	782	GTGACCTGGGCGGAGCCCGGGGCGCTTCCGTTCCCGAGACGACCAAGAGGAGAT	841	/rpl_type=dispersed complement(462..771) /rpl_family="Alu"
Db	780	GTGCACCTGGTGGCGGACCCGCGGCGCTGCTGCGCTCCCGGAGCAGACAGCAGGCT	839	/rpl_type=dispersed 772..945 /rpl_family="L2"
QY	842	CTCATGATGACAGCTGCGCATTTGTATGGGGAGCATGACGACGAATCTCAAGAGAGAC	901	/rpl_type=dispersed 1169..1338 /rpl_family="MIR"
Db	840	CTGGCGCGGTGACAAAGGCATCGTGTGGGACCAACAGGACGCGTGGGTGGAGCCGACCC	899	/rpl_type=dispersed 1169..1338 /rpl_family="MIR"
QY	902	CAACCTACTATGTGATGACAGGTGATGTGCCAAGCAGCTGAGATCTTCAAGACCATC	961	/rpl_type=dispersed 1169..1338 /rpl_family="MIR"
Db	900	GGCGTGGCGGTGGCGCGGAGGTGGCGGTAGCCACCTACGATCGCCGAGCGCGCACAA	959	/rpl_type=dispersed 1169..1338 /rpl_family="MIR"
QY	962	CAGTCTTGCCCGCAGGCGCTGACAGAGAGCTACCTGCTTGCGCATATGAGACCTGAGCT	1021	/rpl_type=dispersed complement(1553..1804) /rpl_family="L153"
Db	960	CTCAGAGCGCGCACCCCTTTCTGCGGGCGCGCTACCGCTGCTGGCGCTTCAGAGACCTGGCG	1019	/rpl_type=dispersed 2616..2916 /rpl_family="Alu"
QY	1022	CGAGCCCGCTGTGGCCGACACTTCCCGATGTATGATTCGTGGGATTGGAATCTTGCCC	1081	/note=.3532 3309..3532 /rpl_type=dispersed
Db	1020	CGGAGCGCGGTGGAGAAATCCGCGCTCTACGCGCTCACTGAGGCTCACTGTCACAGCA	1079	/note=.4028 3865..4028 /note=.similar to EST H12297 (EST cluster Hs.31147)"
QY	1082	CAGTTCAGACCTGGGAGCATATCATGACCCGAGGCAAGGCGCATGGGTTG-----ACCCAC	1135	/rpl_family="Alu"
Db	1080	CAGTTCAGAGGCTGTGATTCATATCATGACCCGAGGATCTGAGCTGTGGCGCGCGGAA	1139	/rpl_type=dispersed 4793..5250 /rpl_family="L2"
QY	1136	GCTTTCACACAAATGCCAGGAGTGCCTTAATGTCTCCAGGCTTGGCGCTGATCTTTG	1195	/rpl_type=dispersed complement(5251..5329) /rpl_family="MIR"
Db	1140	GCTTTCAGAACTGTGTCAGAGAAATGCCCTCACGCTCCAGCGCGCGCGCATGGCGTG	1199	/rpl_type=dispersed complement(5768..6080) /rpl_family="Alu"
QY	1196	CCCTATGAAAAGCTTTCTGCAGCTTCAGAAAGCTGTGGCGATGCCATGATTTGCTGGGC	1255	/rpl_type=dispersed complement(6081..6265) /rpl_family="Alu"
Db	1200	CCCTTTCAGAGATCCGCGCGGTGCAGAACTGTGCCGTGTGGCGTGGCGCTGGGCG	1259	/rpl_type=dispersed 611..6567 /note=.similar to EST AA01593 (EST cluster Hs.6653)"
QY	1256	TACGCGCAGCTGACATCTGAACAAGACAGAAACTGTGCTGATCTTCTTG	1309	/rpl_type=dispersed 8035..8350 /rpl_family="Alu"
Db	1260	TACGCGCGCTGTGATCTGTAGGAGAGACAGCAGCACTGCCCTTATCTGTGGTG	1313	/rpl_type=dispersed 10545..11049 /note=.similar to EST A1950923 (EST cluster Hs.6653)"
RESULT 13				
LOCUS	AF219991	71503 bp	DNA	linear
DEFINITION				PRI 26-OCT-2000
				Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase
				(CHST6) and corneal N-acetylglucosamine-6-O-sulfotransferase
				(CHST6) genes, complete cds.
ACCESSION				
VERSION	AF219991			GI:11023147
KEYWORDS				
SOURCE				
ORGANISM				Homo sapiens.
				Homo sapiens
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
				Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE				
AUTHORS				1 (bases 1 to 71503)
				Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T.,
				Nakamura,T., Dotsu,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
				Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
				Mucular corneal dystrophy type I and type II are caused by distinct
				mutations in a new sulfotransferase gene
TITLE				Nat. Genet. 26 (2), 237-241 (2000)
JOURNAL				
MEDLINE	20472330			
PUBMED	11017086			
REFERENCE				
AUTHORS				2 (bases 1 to 71503)
TITLE				Akama,T.O. and Fukuda,M.N.
JOURNAL				Direct Submission
				Submitted (29-DEC-1999) Glycobiology Program, The Burnham
				Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES				
Source				Location/Qualifiers
				1..71503
				/organism="Homo sapiens"
				/db_xref="taxon:9606"
				/chromosome="16"
				/map="16q22"
				/clone="CITB-483K2"
				358..461
				/rpl_family="L2"
repeat_region				



Db	175	CTTTCACACGACACCCCGACGCTCTTCTACTAATGAGCCCGCGTGCCAGCTGTGGACC	234
QY	422	ACCTTCACAGCAGACACCGCTGTGATGCTGCACATGGCTGTGCGGAGCTCTATACGGACC	481
Db	235	ACCCGTGTGACGAGGAGCGCCGCAACGGTGTGCATATGGTGTGTGGCAACCTGGTCCGCC	294
QY	482	GCTCTTCTGTGCGACATGAGCGCTCTTTGATGCTCTACATGGAACCTGGTGTCCCGGAGACG	541
Db	295	GCTTTCCTGTGCGACATGAGCGTGTGTGATGCTCTATGTCG--CCTTGGCGCGCAACCTG	351
QY	542	TCCAGCCCTTTCAGTGGGAGAACAGACCGCGGCCCGTGTGTGACCTGCGCTGTGACATC	601
Db	352	TCCAGCCTTTCAGTGGGAGAACAGACCGCGGCCCGTGTGTGACCTGCGCTGTGACATC	411
QY	602	ATCCCAAGATGAAATATATCCCGGGCTACATGCGACAGGCTCCTGTGACATCAACACCC	661
Db	412	TTTCCCGAGGCGCATGACGACGAGCGCGGTGTGACAGCAGCTGTGGCGCGGAGTCC	471
QY	662	TTTGAAGTGTGTGAGAAAGAGCGCTCGCCGCTCTACAGACCAGTGTGTCTAAGAGGTGCGC	721
Db	472	TTTACCCCTGGCGCGGAGGCGCTCGCCGCTCTACAGACCAGTGTGTCTAAGAGGTGCGC	531
QY	722	TTCTTCAACCTGACATCTCCTTACCCCGGTGTGAAAGACCCCTCCTCACTGATATC	781
Db	532	TTCTTCAACCTGACATCTCCTTACCCCGGTGTGAAAGACCCCTCCTCACTGATATC	591
QY	782	GTCGACCTGTGTCGGGAGACCCCGGGCGCTGTTCGCTTCCGAGAACGACAAAGGAGAT	841
Db	592	GTCGACCTGTGTCGGGAGACCCCGGGCGCTGTTCGCTTCCGAGAACGACAAAGGAGAT	651
QY	842	CTCATGATTGACATGCGATGTGTATGGGCGACGATGAGCAAGAACTCAAGAGAGAGAC	901
Db	652	CTGCGCGTGTGACAAAGCGATGCTGTGGGCAACCAAGCGACGTAAGAGCGGACCC	711
QY	902	CAACCCCTGATGTGATGATCAGATCTATGCGCAAAAGCCAGTGTGAGATATACAGACATC	961
Db	712	GCCCTGGCGCTGTGTGCGGAGGTGTGTGCGGTAGCCACGTACCCATGCGCGAGCGCCACA	771
QY	962	CAGTCTTGGCCCAAGGCCCTGTGAGAGACGTACCTGCTTGTGCGTATGAGACCTGGCT	1021
Db	772	CTCAAGCGCCACACCTTCTGTGCGGCGGTACACCGCTGTGCGGCTTGAAGACCTGGCG	831
QY	1022	CGAGCCCTGTGCGCCACACTTCCGAGTGTATGATTCGTGGGATGTGATTTCTTGCC	1081
Db	832	CGGCGCGCTGTGCGCAATCCGTGCGCTCTACGCTTTCACCTGGGCTAGTCTACGCGCA	891
QY	1082	CATCTTCGACCTGGGTGCATATCATCATCCCGAGGCAAGGCGCATGGGAGT-----ACCAC	1135
Db	892	CAGCTGAGCGCTGTGATCATATACATCACCCACGAGATGTGAGCTGTGGCGCGCGGCAA	951
QY	1136	GCTTTCACACAAATGCGCAGGAGTGCCTTATGTCTCCCAAGGCTTGGCGCTGTGCTT	1195
Db	952	GCTTTCACACAAATGCGCAGGAGTGCCTTATGTCTCCCAAGGCTTGGCGCTGTGCTT	1011
QY	1196	CCCTATGAAAGGTTTCTGCACTTASAAAGCCTGTGCGATGCGCATATTTGCTGGC	1255
Db	1012	CCCTTGTGCAGATGCGCGCGGTGTGASGAACATGTGCGCTGTGGCTGACAGTGTGGC	1071
QY	1256	TACGCCACGTCAGATCTGAAACAGACAGAAACCTGTTGCTGTGATCTTCTG	1309
Db	1072	TACGCCCTGTGTACTGTAGAGACAGAGGCAACCTGCTGCTGTGATCTTCTG	1125

FEATURE 11	
AF219990	
LOCUS	2544 bp mRNA linear PRI 26-OCT-2000
DEFINITION	Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase
	(CHST6) mRNA, complete cds.
ACCESSION	AF219990
VERSION	AF219990.1 GI:11023145
KEYWORDS	
SOURCE	Homo sapiens.

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	TITLE	JOURNAL	FEATURES
1	(bases 1 to 2544)					2	(bases 1 to 2544)		
1	Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T., Nakamura,T., Dotsu,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Ozaki,K., Kinoshita,S., Shimomura,Y., Taniguchi,A. and Fukuda,M.N.	Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene	Nat. Genet. 26 (2), 237-241 (2000)	20472330	11017086	2	(bases 1 to 2544)		
2	Akama,T.O. and Fukuda,M.N.	Direct Submission	Submitted (29-DEC-1999) Glycobiology Program, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA			1	.2544		
3						1	.2544		
4						2	(bases 1 to 2544)		
5						3			
6						4			
7						5			
8						6			
9						7			
10						8			
11						9			
12						10			
13						11			
14						12			
15						13			
16						14			
17						15			
18						16			
19						17			
20						18			
21						19			
22						20			
23						21			
24						22			
25						23			
26						24			
27						25			
28						26			
29						27			
30						28			
31						29			
32						30			
33						31			
34						32			
35						33			
36						34			
37						35			
38						36			
39						37			
40						38			
41						39			
42						40			
43						41			
44						42			
45						43			
46						44			
47						45			
48						46			
49						47			
50						48			
51						49			
52						50			
53						51			
54						52			
55						53			
56						54			
57						55			
58						56			
59						57	</		

QY 765 CCTCAACCTGCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 824  
D 985 CCTCAACCTGCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1044  
QY 825 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 884  
D 1045 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1104  
QY 885 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 944  
D 1105 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1164  
QY 945 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1004  
D 1165 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1224  
QY 1005 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1064  
D 1225 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1284  
QY 1065 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1124  
D 1285 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1344  
QY 1125 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1184  
D 1345 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1404  
QY 1185 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1244  
D 1405 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1464  
QY 1245 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1304  
D 1465 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1524  
QY 1305 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1364  
D 1525 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1584  
QY 1365 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1424  
D 1585 AACGACAAAGGAGATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1644  
RESULT 9  
AX381256 517 bp DNA linear PAT 18-MAR-2002  
LOCUS AX381256  
DEFINITION Sequence 194 from Patent WO0212280.  
ACCESSION AX381256  
VERSION AX381256.1 GI:19576075  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Pyle, R.A., Xu, J., and Secrist, H.  
TITLE Compositions and methods for the therapy and diagnosis of colon  
cancer  
JOURNAL WO 0212280-A 194 14-FEB-2002;  
CORINA CORPORATION (US)  
FEATURES  
Location/Qualifiers  
1. 517  
BASE COUNT 121 a 130 c 124 g 142 t  
ORIGIN  
Query Match 24.9%; Score 506.4; DB 6; Length 517;  
Best Local Similarity 99.8%; Pred. No. 8,9e-120;  
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1352 GAAGGCTTTCGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1411  
D 10 GAAGGCTTTCGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1464  
QY 1412 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1471  
D 70 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1524  
QY 1472 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1531  
D 130 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1584  
QY 1532 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1591  
D 190 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1644  
QY 1592 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1651  
D 250 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1704  
QY 1652 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1711  
D 310 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1764  
QY 1712 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1771  
D 370 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1824  
QY 1772 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1831  
D 430 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1884  
QY 1832 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1891  
D 490 TGCCATATCTGTCGACCTGTGTCGGGAGACCCCGGCGCTTTCGGTCCCGAG 1944  
RESULT 10  
AX327330 1647 bp DNA linear PAT 07-JAN-2002  
LOCUS AX327330  
DEFINITION Sequence 15 from Patent WO0179468.  
ACCESSION AX327330  
VERSION AX327330.1 GI:18097876  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Policky, J.L., Hatfield, A., Burford, N., Ring, H.Z., Lal, P.,  
Tribouley, C.M., Yao, M.G., Yue, H., Tang, Y.T., Patterson, C., Das, D.,  
Sanjanwala, M.S., Gandhi, A.R., Reddy, R., Khan, F.A., Baughn, M.R.,  
Ramkumar, J., Griffin, J.A., and Au-Young, J.  
TITLE Drug metabolizing enzymes  
JOURNAL Patent: WO 0179468-A 15 25-OCT-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
Location/Qualifiers  
1. 1647  
BASE COUNT 287 a 560 c 501 g 299 t  
ORIGIN  
Query Match 20.4%; Score 414.4; DB 6; Length 1647;  
Best Local Similarity 64.5%; Pred. No. 6e-96;  
Matches 654; Conservative 0; Mismatches 351; Indels 9; Gaps 2;







QY	155	TCTTGCAGAGGCTTCCACTTTCAGCACATGCTACTGCTATAAAAAATGAAGCTCTGCTG	214
QY	1775	AGTCCCGAGTGATTCACGAGGAAGAGAGTGGGACACAGGTGGATTCCTACTATAGCTT	1834
Db	144198	AGTTCCTCCAGTGGATTCACGAGGAAGAGTGGGACACAGGTGGATTCCTACTATAGAGCTT	144257
QY	1835	GACCATTCACAGCTATCGGTTAATCAGAAATATATAACAAATCTCTGCACAAAGAGCAAG	1894
Db	144258	GACCATTCACAGCTATCGGTTAATCAGAAATATATAACAAATCTCTGCACAAAGAGCAAG	144317
QY	1895	CTCTTAAGTTCACAGAGGTGCTGGGCTGCATTGAAATACCTTCCCTCTGCATTTTCC	1954
Db	144318	CTCTTAAGTTCACAGAGGTGCTGGGCTGCATTGAAATACCTTCCCTCTGCATTTTCC	144377
QY	1955	CATCATCATAAGAGACTTTCAGCTGTGAAGCTGCCATCTGTTATACATAAATTCACCAAT	2014
Db	144378	CATCATCATAAGAGACTTTCAGCTGTGAAGCTGCCATCTGTTATACATAAATTCACCAAT	144437
QY	2015	AAG 2017	
Db	144438	AAG 144440	
RESULT 6			
LOCUS	AF149783	1333 bp	mRNA
DEFINITION	Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.		
ACCESSION	AF149783		
VERSION	AF149783.1	GI:13897503	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
	Euhariyola, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1333)		
AUTHORS	Yeh,J.C., Hiraoka,N., Petryniak,B., Nakayama,J., Elies,L.G.,		
	Rabuk,D., Hirsengau,O., March,J.D., Loe,J.B. and Fukuda,M.,		
	Novel sulfated lymphocyte homing receptors and their control by a		
	Core1 extension beta 1,3-N-acetylglucosaminyltransferase		
	Cell 105 (??), 957-969 (2001)		
TITLE			
JOURNAL			
MEDLINE	21332592		
PUBMED	11439191		
REFERENCE	2 (bases 1 to 1333)		
AUTHORS	Hiraoka,N. and Fukuda,M.		
TITLE	Direct Submission.		
JOURNAL	Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901		
	North Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES			
source	1..1333		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	93..1253		
	/note="LST"		
	/codon_start=1		
	/product="L-selectin ligand sulfotransferase"		
	/protein_id="AAK48417.1"		
	/db_xref="GI:13897504"		
	/translation="MLPKKKKLLFLYSQALLALFFHNSHISLSKKAQPERHHH		
	VLIVSSWRSGSEVGLGQHPDVFYLMPEAHVWMTFKOSTAMLMHAYDLIRAVF		
	LCQDSVDAYKPEGRKRSLSLEQWNSRALCSAPACDIIPDELTIPAHCSLCSQDP		
	FEVWEKACRSYSHVILKLEVFREFNQSILPLKDPSLNHIVHLVDRPAAVRSRETRK		
	GDMLADIRIVAGQHEOKLKEDQPYVMQVQCSQLETKIQLSPRALQRYLLVRY		
	EDLARVAVQTSRMVEFVGLLEFLHLDQVWNTIRGKGMDPAHHTNARADLVNSQAM		
	RMSLPYKVSLSLQKACDQANMLLGRVRSRQEPKRLDLILSLTWYPEQIH"		
BASE COUNT	286 a 393 c 336 g 318 t		
ORIGIN			
Query Match	62.1%, Score 1261.6; DB 9; Length 1333;		
Best Local Similarity	99.7%; Pred. No. 0;		
Matches 1264; Conservative	0; Mismatches 4; Indels 0; Gaps 0;		

Db	66	TCCTTAAGGCTCCACTCCACTCCAGACAAAGCTACTACCTTAATAAAATGAAGTCCGTG	125
QY	215	TTTCTGCTTTCOCAGATGGCCATCTTGCTCTATTTCTCCATGTAAAGCCAACTC	274
Db	126	TTTCTGCTTTCOCAGATGGCCATCTTGCTCTATTTCTCCATGTAAAGCCAACTC	185
QY	275	AGCTCCCTGTCTATGAAGGACACAGCCGAGAGCATGCAGTCTGTCTGTCTCCGG	334
Db	186	AGCTCCCTGTCTATGAAGGACACAGCCGAGAGCATGCAGTCTGTCTGTCTCCGG	245
QY	335	CGCTCTGCTCTTCTTTTGTGGGGACGTTTTTGGCAGCACCCAGATGTTTTCTACCTG	394
Db	246	CGCTCTGCTCTTCTTTTGTGGGGACGTTTTTGGCAGCACCCAGATGTTTTCTACCTG	305
QY	395	ATGAGAGCCCGCTGGCACGCTGTGGATGACCTTAAACAGAGACCGCCGTGGATGTGAC	454
Db	306	ATGAGAGCCCGCTGGCACGCTGTGGATGACCTTAAACAGAGACCGCCGTGGATGTGAC	365
QY	455	ATGGCTGTGGGGATCTGATACGGCCGCTTCTTGTGGACATGAGCGTCTTGTATGCC	514
Db	366	ATGGCTGTGGGGATCTGATACGGCCGCTTCTTGTGGACATGAGCGTCTTGTATGCC	425
QY	515	TACATGGAACCTGCTCCCCGGAGACAGTCCAGCTCTTTCACTGGGAGAAACAGCCGGGCC	574
Db	426	TACATGGAACCTGCTCCCCGGAGACAGTCCAGCTCTTTCACTGGGAGAAACAGCCGGGCC	485
QY	575	CTGTGTTCTGCACCTGCGCTGTGACATCATCCCAAGATGAATATATCCCCGGGCTAC	634
Db	486	CTGTGTTCTGCACCTGCGCTGTGACATCATCCCAAGATGAATATATCCCCGGGCTAC	545
QY	635	TGCAGGCTCTGTGCAGTCAACAGCCCTTTAGCTGTGGAGAGAGCCTGCCGCTCTAC	694
Db	546	TGCAGGCTCTGTGCAGTCAACAGCCCTTTAGCTGTGGAGAGAGCCTGCCGCTCTAC	605
QY	695	AGCCACGCTGTGCTCAAGGAGGTGGCTTTTAACTCAGTCCCTCTACCCGCTGGTG	754
Db	606	AGCCACGCTGTGCTCAAGGAGGTGGCTTTTAACTCAGTCCCTCTACCCGCTGGTG	665
QY	755	AAAGACCCCTCCCTCAACCTGCATATCTGATGACAGCTGGTCCGGAGACCCCGGGCCGTGTC	814
Db	666	AAAGACCCCTCCCTCAACCTGCATATCTGATGACAGCTGGTCCGGAGACCCCGGGCCGTGTC	725
QY	815	CGTTCCCGGAGAAACCCAAAGGAGATCTCATGATTGAGATGCTGATTTGTATGGGAGAG	874
Db	726	CGTTCCCGGAGAAACCCAAAGGAGATCTCATGATTGAGATGCTGATTTGTATGGGAGAG	785
QY	875	CATAGACAGAAACTCAAGAGAGAGAGCAACCTTACTATGATGCAGGTCTATCGCCAA	934
Db	786	CATAGACAGAAACTCAAGAGAGAGAGCAACCTTACTATGATGCAGGTCTATCGCCAA	845
QY	935	AGCCAGCTGCAGATCTCAAGACCATTCAGTCTTGCCCAAGGCCCTGCAGGACGCTAC	994
Db	846	AGCCAGCTGCAGATCTCAAGACCATTCAGTCTTGCCCAAGGCCCTGCAGGACGCTAC	905
QY	995	CTGCTGTGGGCTTATGAGAGCTGGCTGCAGAGCCCTGTGGCCAGAGTCCCGAATGTAT	1054
Db	906	CTGCTGTGGGCTTATGAGAGCTGGCTGCAGAGCCCTGTGGCCAGAGTCCCGAATGTAT	965
QY	1055	GAATTCGTGGGATTTGAATTTCTTGCCCATCTTCAGACCTGGGATGCATTAATCAACCGCA	1114
Db	966	GAATTCGTGGGATTTGAATTTCTTGCCCATCTTCAGACCTGGGATGCATTAATCAACCGCA	1025
QY	1115	GGCAAGGGCATGGGTGACACACCTTTCACAAACAAAGCCAGGAGATGGCCTTATGTCTCC	1174
Db	1026	GGCAAGGGCATGGGTGACACACCTTTCACAAACAAAGCCAGGAGATGGCCTTATGTCTCC	1085
QY	1175	CAGGCTGTGGCGCTGCTTTTGGCCCTTATGAAAAGTTTCTGCATTCAGAAAACCTGTGCG	1234
Db	1086	CAGGCTGTGGCGCTGCTTTTGGCCCTTATGAAAAGTTTCTGCATTCAGAAAACCTGTGCG	1145
QY	1235	GATGCCATGAATTTGCTGGGCTACCGCCACAGTCAGATCTTGAAACAAAGAACAGAAACTG	1294
Db	1146	GATGCCATGAATTTGCTGGGCTACCGCCACAGTCAGATCTTGAAACAAAGAACAGAAACTG	1205

REFERENCE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 183228)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Sep 29, 2001 this sequence version replaced gi:14589436.  
COMMENT Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

FEATURES  
Location/Qualifiers  
1. 183228  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-510W2"

BASE COUNT 51901 a 42098 c 40026 g 49203 t

ORIGIN

Query Match 91.4% Score 1856.6 DB 9: Length 183228:  
Best Local Similarity 95.8% Pred. No. 0:  
Matches 1859: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 155 TCTTCAAGGCTCTCACTACGACATGCTACGCTAAATAAATGAAGCTCTGCTG 214  
DB 142578 TCCCTAAAGGCTCTCACTACGACATGCTACGCTAAATAAATGAAGCTCTGCTG 142637  
QY 215 TTCTGCTTCCAGATGCGCATCTGGCTCTATTCTTCACATGATACGACCAACATC 274  
DB 142638 TTCTGCTTCCAGATGCGCATCTGGCTCTATTCTTCACATGATACGACCAACATC 142697  
QY 275 AGCTCCCTGTATGAAGGACACAGCCGAGGCGATGCGCTGGTCTGCTCTGCG 334  
DB 142698 AGCTCCCTGTATGAAGGACACAGCCGAGGCGATGCGCTGGTCTGCTCTGCG 142757  
QY 335 CGCTCTGCTCTTTTGTGGGGAGCTTTTGGGGAGACCCAGATGTTTCTACCTG 394  
DB 142758 CGCTCTGCTCTTTTGTGGGGAGCTTTTGGGGAGACCCAGATGTTTCTACCTG 142817  
QY 395 ATGGAGCCGCGCTGGACGCTGTGATGACCTTCAAGAGAGACCGCTGTGATGCTGAC 454  
DB 142818 ATGGAGCCGCGCTGGACGCTGTGATGACCTTCAAGAGAGACCGCTGTGATGCTGAC 142877  
QY 455 ATGGAGCTGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTATGCC 514  
DB 142878 ATGGAGCTGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTATGCC 142937  
QY 515 TACATGGAACCTGCTCCCGGAGAGATCCAGCTCTTTCAGTGGAGAACAGCCGGGCG 574  
DB 142938 TACATGGAACCTGCTCCCGGAGAGATCCAGCTCTTTCAGTGGAGAACAGCCGGGCG 142997  
QY 575 CTGTGTTGTGACCTGCTGTGACATCATCCCAAGATGAATCAATCCCGGGGCTCAC 634  
DB 142998 CTGTGTTGTGACCTGCTGTGACATCATCCCAAGATGAATCAATCCCGGGGCTCAC 143057  
QY 635 TGCAGGCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAAAGGCGCTGCCCTCTAC 694

DB 143058 TGCAGGCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAAAGGCGCTGCCCTCTAC 143117  
QY 695 AGCCAGCTGTGCTCAAGAGAGTGCGCTTCTTCACTGACCTGCTCTACCCGCTGCTG 754  
DB 143118 AGCCAGCTGTGCTCAAGAGAGTGCGCTTCTTCACTGACCTGCTCTACCCGCTGCTG 143177  
QY 755 AAAGACCCCTCCCTCAACCGCATATGCGACCTGCTGCGGAGACCCCGGGCGGCTGC 814  
DB 143178 AAAGACCCCTCCCTCAACCGCATATGCGACCTGCTGCGGAGACCCCGGGCGGCTGC 143237  
QY 815 CGTTCGCGAGAACGACCAAGAGGAGATCTCATGATTTGATGACAGTGTGATGAGGCGAG 874  
DB 143238 CGTTCGCGAGAACGACCAAGAGGAGATCTCATGATTTGATGACAGTGTGATGAGGCGAG 143297  
QY 875 CATGAGAGAAATCTCAAGAGAGAGAGACCAACCTACTATGTTGATGAGGCTGCGCA 934  
DB 143298 CATGAGAGAAATCTCAAGAGAGAGAGACCAACCTACTATGTTGATGAGGCTGCGCA 143357  
QY 935 AGCCAGCTGTGAGATCTCAAGAGACCATCACTGCTTCCGAGGCGCTGAGAACGCTAC 994  
DB 143358 AGCCAGCTGTGAGATCTCAAGAGACCATCACTGCTTCCGAGGCGCTGAGAACGCTAC 143417  
QY 995 CTGCTTGTGCGCTATGAGACCTGCTGAGAGCCCTGTGCGGACACTTCCGATGATAT 1054  
DB 143418 CTGCTTGTGCGCTATGAGACCTGCTGAGAGCCCTGTGCGGACACTTCCGATGATAT 143477  
QY 1055 GAATTCGTGGGATTTGGAATCTTGGCCCATCTTCAACAGCGGGTGTGATACATCCCA 1114  
DB 143478 GAATTCGTGGGATTTGGAATCTTGGCCCATCTTCAACAGCGGGTGTGATACATCCCA 143537  
QY 1115 GGCAGAGGCGATGGGATGACACGCTTTCCACACAATATGCGAGGATGCTTAATGCTCC 1174  
DB 143538 GGCAGAGGCGATGGGATGACACGCTTTCCACACAATATGCGAGGATGCTTAATGCTCC 143597  
QY 1175 CAGGCTTGGCGCTGCTGCTTGGCCCTATGAAAGGTTCTGCGATGAGAAAGCTGTGCG 1234  
DB 143598 CAGGCTTGGCGCTGCTGCTTGGCCCTATGAAAGGTTCTGCGATGAGAAAGCTGTGCG 143657  
QY 1235 GATGCCATGAAATTTGCTGGGCTACCGCCAGCTGATGATGTAACAGAGAGAAACCTG 1294  
DB 143658 GATGCCATGAAATTTGCTGGGCTACCGCCAGCTGATGATGTAACAGAGAGAAACCTG 143717  
QY 1295 TTGCTGATTTCTGTCTACCTGACCTGCTGCTGACCAATCACTAAGAGGTTAGAA 1354  
DB 143718 TTGCTGATTTCTGTCTACCTGACCTGCTGCTGACCAATCACTAAGAGGTTAGAA 143777  
QY 1355 GGCCTTGTGCTGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1414  
DB 143778 GGCCTTGTGCTGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143837  
QY 1415 CTACATCTGTGAGCTTAACATGATGCTGTGAGATCACTGATGATGATGCTGCTGCTG 1474  
DB 143838 CTACATCTGTGAGCTTAACATGATGCTGTGAGATCACTGATGATGATGCTGCTGCTG 143897  
QY 1475 ACAAGTGTCAAGCAGAGAGACTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1534  
DB 143898 ACAAGTGTCAAGCAGAGAGACTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143957  
QY 1535 AACCTATGTGAGCAGACATCCACAGTGAAGAGAGGATTTGCTGCTGCTGCTGCTGCTG 1594  
DB 143958 AACCTATGTGAGCAGACATCCACAGTGAAGAGAGGATTTGCTGCTGCTGCTGCTGCTG 144017  
QY 1595 GATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1654  
DB 144018 GATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 144077  
QY 1655 ACAATATCAATGATGATGATCAATAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1714  
DB 144078 ACAATATCAATGATGATGATCAATAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 144137  
QY 1715 GATCTTCAACCAAGAGCTCACAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGG 1774

D8 AAGCCAGGCGACAGAGTCTTCCACTTCAGACATGCTACGCTAATAAATAAGAGCT 147  
OY 208 CTTCTCTTTCTGTGTTCACAGATGGCCATCTGGCTCTATTCTTCACATGTACAGCCA 267  
Db 148 CTTCTCTTTCTGTGTTCACAGATGGCCATCTGGCTCTATTCTTCACATGTACAGCCA 207  
OY 268 CAACATCAGCTCCCTGTCTATGAAGGACAGCCGAGCCGATCAGATGCTGTCTGTG 327  
Db 208 CAATATAGCTCCCTGTCTATGAAGGACAGCCGAGCCGATCAGATGCTGTCTGTG 267  
OY 328 TTCTGTGCGCTCTGTGTCTTTTGTGGGCGAGCTTTTGGGCGAGCCAGATGTTTT 387  
Db 268 TTCTGTGCGCTCTGTGTCTTTTGTGGGCGAGCTTTTGGGCGAGCCAGATGTTTT 327  
OY 388 CTACCTATGAGAGCCGCTGGCAGCTGTGATGACCTTGAAGCAGAGCAGCCGCTGAT 447  
Db 328 CTACCTATGAGAGCCGCTGGCAGCTGTGATGACCTTGAAGCAGAGCAGCCGCTGAT 387  
OY 448 GCTGACATGCTGTGTGGGATCTGATAGGGCGCTCTTCTTGTGGACATGAGCGCTTT 507  
Db 388 GCTGACATGCTGTGTGGGATCTGATAGGGCGCTCTTCTTGTGGACATGAGCGCTTT 447  
OY 508 TGATGCTATGATGAGACCTGTGCTGCGGAGACAGCTGACGCTCTTTCAGTGGGAGAGAG 567  
Db 448 TGATGCTATGATGAGACCTGTGCTGCGGAGACAGCTGACGCTCTTTCAGTGGGAGAGAG 507  
OY 568 CCGGGCCCTGTGTGTGACCTGCTGCTGTGACATCATCCCAAGATGAAATCATCCCGG 627  
Db 508 CCGGGCCCTGTGTGTGACCTGCTGCTGTGACATCATCCCAAGATGAAATCATCCCGG 567  
OY 628 GGCCTACATGAGAGCTCTGTGTGACATCAAGAGCCCTTTGAGCTGTGTGAGAGAGCCCTGCGG 687  
Db 568 GGCCTACATGAGAGCTCTGTGTGACATCAAGAGCCCTTTGAGCTGTGTGAGAGAGCCCTGCGG 627  
OY 688 CTCTACAGCAGCTGTGTGTGACAGAGAGTGTGCTTCAACCTGAGAGCCCTCTACCC 747  
Db 628 CTCTACAGCAGCTGTGTGTGACAGAGAGTGTGCTTCAACCTGAGAGCCCTCTACCC 687  
OY 748 GCTGTGAAAGAGCCCTCTCTCAACCTGATATGCTGACCTGTGCTGGGAGCCCGGGG 807  
Db 688 GCTGTGAAAGAGCCCTCTCTCAACCTGATATGCTGACCTGTGCTGGGAGCCCGGGG 747  
OY 808 CGTGTTCGCTCCGAGAGAGCAGCAAGAGAGATCTCATATTGACAGTGTGAT 867  
Db 748 CGTGTTCGCTCCGAGAGAGCAGCAAGAGAGATCTCATATTGACAGTGTGAT 807  
OY 868 GGGGAGAGATGAGCAAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927  
Db 808 GGGGAGAGATGAGCAAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867  
OY 928 CTGCAAGAGAGAGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 987  
Db 868 CTGCAAGAGAGAGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927  
OY 988 AGCTTACCTGCTGTGTGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1047  
Db 928 AGCTTACCTGCTGTGTGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 987  
OY 1048 AATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107  
Db 988 AATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1047  
OY 1108 CACCCAG 1167  
Db 1048 CACCCAG 1107  
OY 1168 TGTCTCCAGAGCTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1227  
Db 1108 TGTCTCCAGAGCTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1167  
OY 1228 CTGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287  
Db 1168 CTGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1227

OY 1288 AAGCCAGGCGACAGAGTCTTCCACTTCAGACATGCTACGCTAATAAATAAGAGCT 147  
Db 1228 AAGCCAGGCGACAGAGTCTTCCACTTCAGACATGCTACGCTAATAAATAAGAGCT 147  
OY 1348 TTGAGAGAGCTTTGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407  
Db 1288 TTGAGAGAGCTTTGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407  
OY 1408 GCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1467  
Db 1348 GCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1407  
OY 1468 TGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1527  
Db 1408 TGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1467  
OY 1528 ACTGGGAGAGCTTTATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1587  
Db 1468 ACTGGGAGAGCTTTATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527  
OY 1588 TTTTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1647  
Db 1528 TTTTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1587  
OY 1648 GAGAGAGAGAGATGAG 1707  
Db 1588 GAGAGAGAGAGATGAG 1647  
OY 1708 GGGAGAGAGAGCTTTTCCAG 1767  
Db 1648 GGGAGAGAGAGCTTTTCCAG 1707  
OY 1768 CCCTTGTGAGTCTTCCAG 1827  
Db 1708 CCCTTGTGAGTCTTCCAG 1767  
OY 1828 TGAGCTGTGAG 1887  
Db 1768 TGAGCTGTGAG 1827  
OY 1888 GAGAGAGAGCTTTTATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1947  
Db 1828 GAGAGAGAGCTTTTATGTGAG 1887  
OY 1948 ATTTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2007  
Db 1888 ATTTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1947  
OY 2008 CCCAATAG 2032  
Db 1948 CCCAATAG 2008

RESULT 5  
AC010547 183228 bp DNA linear PRI 29-SEP-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.  
DEFINITION AC010547  
ACCESSION AC010547.9 GI:15808510  
VERSION  
KEYWORDS HTG.  
SOURCE  
ORGANISM Homo sapiens.  
Homosapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 183228)  
2 (bases 1 to 183228)  
AUTHORS DOE Joint Genome Institute.  
TITLE Unpublished  
JOURNAL  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint



Db	1321	TGTCCCTGAGCAATTCACCTAAGAGGGCTTTGAGAAGGCTTTGCTGCCACCTGCTGTCAGCC	1380
Qy	1381	TCAGTCACCTTTCTCGAATGCTTCTGAGCCTTCCCTACATCTCTGTAGCCTTAACATCATG	1440
Db	1381	TCAGTCACCTTTCTCGAATGCTTCTGAGCCTTCCCTACATCTCTGTAGCCTTAACATCATG	1440
Qy	1441	TCTGTGGATATCACACTAGAGTGTGAGTTGTGTCCACAGCTGCTCAAGCAGGAAGGACTTTT	1500
Db	1441	TCTGTGGATATCACACTAGAGTGTGAGTTGTGTCCACAGCTGCTCAAGCAGGAAGGACTTTT	1500
Qy	1501	GTGTCCATGCTTGTGTCTAGAAACAGATGSGGGAACCTTATGTAGAGCAGCATCTCCAC	1560
Db	1501	GTGTCCATGCTTGTGTCTAGAAACAGATGSGGGAACCTTATGTAGAGCAGCATCTCCAC	1560
Qy	1561	CAGTGAACAGGGTATGTCTCTTTCTTTCTTTCTTTGATCTTCTCTGTCTGGGCAAGCTTCAG	1620
Db	1561	CAGTGAACAGGGTATGTCTCTTTCTTTCTTTCTTTGATCTTCTCTGTCTGGGCAAGCTTCAG	1620
Qy	1621	ACACTTTGTGGCCTGGAGGACCTATTAAAGCAGCAGCAGATAGTGGATTTGATTCATTA	1680
Db	1621	ACACTTTGTGGCCTGGAGGACCTATTAAAGCAGCAGCAGATAGTGGATTTGATTCATTA	1680
Qy	1681	ACCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTTTCACCAAAAGAGCTCACACAGC	1740
Db	1681	ACCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTTTCACCAAAAGAGCTCACACAGC	1740
Qy	1741	ATTTTCCACAGAGATGGAATTTGTAGCCCTTGAGAGTTCCAGTGATTTCAAGGAAGGA	1800
Db	1741	ATTTTCCACAGAGATGGAATTTGTAGCCCTTGAGAGTTCCAGTGATTTCAAGGAAGGA	1800
Qy	1801	GTGGAGACAGGTTGGATGCCCTACTTATGAGTTCATACCATCAGACCTATCGGTAATCAGA	1860
Db	1801	GTGGAGACAGGTTGGATGCCCTACTTATGAGTTCATACCATCAGACCTATCGGTAATCAGA	1860
Qy	1861	AATATGAACAAATATCTCTGCACAAAGACAGCTTTAAGTTACAGAGGTGCTCGGCT	1920
Db	1861	AATATGAACAAATATCTCTGCACAAAGACAGCTTTAAGTTACAGAGGTGCTCGGCT	1920
Qy	1921	TGCATTGAATATCACTTCCCTCCCTGCAATTTCCCATCATAGAAAGATTTGACCTGTG	1980
Db	1921	TGCATTGAATATCACTTCCCTCCCTGCAATTTCCCATCATAGAAAGATTTGACCTGTG	1980
Qy	1981	AAGCTGCCATCTGTTAATPCTTAAATTTCCCAATTAAGAAAAAATTTTTTTT	2032
Db	1981	AAGCTGCCATCTGTTAATPCTTAAATTTCCCAATTAAGAAAAAATTTTTTTT	2032
RESULT 3			
AK026635			
LOCUS	AK026635	2011 bp	linear
DEFINITION	Homo sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to AF131335 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.		
ACCESSION	AK026635	GI:10439531	
VERSION	AK026635.1		
KEYWORDS	Cligo capping; fis (full insect sequence).		
SOURCE	Homo sapiens skin-erling cell carcinoma cell_line:KATO III CDNA to mRNA, clone_lib:KAT clone:KAT11454.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotic Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2011)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human		

Genome Center. Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

NEO human cdna sequencing project supported by Ministry of International Trade and Industry of Japan: cdna full insert sequencing: Research Association for Biotechnology: cDNA library construction: 5'-5,3'-end one pass sequencing: Department of Vitrology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES  
 source  
 Location/Qualifiers  
 1..2011  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="KAT11454"  
 /cell\_line="KATO III"  
 /cell\_type="signet-ring cell carcinoma"  
 /clone\_lib="KAT"  
 /note="Cloning vector pME85FL3"  
 1..2011  
 /note="highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA"

BASE COUNT 482 a 541 c 475 g 513 t

ORIGIN

Query Match 93.8%; Score 1905.4; DB 9; Length 2011;  
 Best Local Similarity 99.9%; Pred. Nid. 0;  
 Matches 1906; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY 111 GTTACTTTACAGGCTTCCTGGGAGGAGGCTTTCACAGCCCGCTTCACAGGCTTCC 170  
 Db 1 GTTACTTTACAGGCTTCCTGGGAGGAGGCTTTCACAGCCCGCTTCACAGGCTTCC 60

QY 171 ACTTCAGACAATGCTACTGCTCTAATAAAAAATGAAGCTCCGCTGTTTCAGTTTCCAGA 230  
 Db 61 ACTTCAGACAATGCTACTGCTCTAATAAAAAATGAAGCTCCGCTGTTTCAGTTTCCAGA 120

QY 231 TGGGCATCTGGCTCTAATCTTCACATGTACAGCACAAACATACGCTCCCTGATATA 290  
 Db 121 TGGGCATCTGGCTCTAATCTTCACATGTACAGCACAAACATACGCTCCCTGATATA 180

QY 291 AGGCACAGCCGAGGCGCATGCAAGGCTGGTCTGTCTCTCTGCTCGGGCTCTGGCTCTTT 350  
 Db 181 AGGCACAGCCGAGGCGCATGCAAGGCTGGTCTGTCTCTCTGCTCGGGCTCTGGCTCTTT 240

QY 351 TTGTGGGGAGGTTTTTGGGACACCCAGATGTTTTCTACTGATGGAGCCGCGCTGGC 410  
 Db 241 TTGTGGGGAGGTTTTTGGGACACCCAGATGTTTTCTACTGATGGAGCCGCGCTGGC 300

QY 411 ACGTGTGATGACCTTCAAGCAGACACCGCGCTGATGCTGCACATGGGTGTGCGGATC 470  
 Db 301 ACGTGTGATGACCTTCAAGCAGACACCGCGCTGATGCTGCACATGGGTGTGCGGATC 360

QY 471 TGAATCGGGGCGCTCTCTTGTGCGACATGAGGCTTTGATGCCATACATGAACTGGTC 530  
 Db 361 TGAATCGGGGCGCTCTCTTGTGCGACATGAGGCTTTGATGCCATACATGAACTGGTC 420

QY 531 CCCGAGACAGTGCAGGCTCTTTCAGTGGGAGAACAGCGGGCCCTGTTGTGCACCTG 590  
 Db 421 CCCGAGACAGTGCAGGCTCTTTCAGTGGGAGAACAGCGGGCCCTGTTGTGCACCTG 480

QY 591 CCTGTGATATATCCCAAGAATGAATATATATATATATATATATATATATATATATAT 650  
 Db 481 CCTGTGATATATCCCAAGAATGAATATATATATATATATATATATATATATATATAT 540

QY 651 GTCAACAGCCCTTTGAGGTGTGTGGAAGAGGCTCCGCGCTCTACAGCCACATGTGTGCTA 710  
 Db 541 GTCAACAGCCCTTTGAGGTGTGTGGAAGAGGCTCCGCGCTCTACAGCCACATGTGTGCTA 600

QY 711 AGGAGGTGGCTCTTTCACCTACGCTGCTCTTACCGCTGTGTGAAGAACCCTCCCTCA 770  
 Db 601 AGGAGGTGGCTCTTTCACCTACGCTGCTCTTACCGCTGTGTGAAGAACCCTCCCTCA 660



FEATURES Location/Qualifiers  
Source 1..2032 /organism="unknown"  
BASE COUNT 468 a 569 c 490 g 505 t  
ORIGIN  
Query Match 100.0%; Score 2032; Db 6; Length 2032;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGTCGAGGCCAGSAGTGCCTCAGTCTGGGGGAAATGCTTCCTCATTTGCTCTCCAG 60  
Db 1 GAGTCGAGGCCAGAGTGCCTCAGTCTGGGGGAAATGCTTCCTCATTTGCTCTCCAG 60  
QY 61 CCCACCTCAGACAGTCTCCCAACCCTTTGAGTCTCAGACAGTGTAAAGCTGTACTTTCA 120  
Db 61 CCCACCTCAGACAGTCTCCCAACCCTTTGAGTCTCAGACAGTGTAAAGCTGTACTTTCA 120  
QY 121 CAGCTCTCGGAGCGAGTGTCTCTCAAGCCCTCTTGCAGAGTCTTCCACTTCAGCAC 180  
Db 121 CAGCTCTCGGAGCGAGTGTCTCTCAAGCCCTCTTGCAGAGTCTTCCACTTCAGCAC 180  
QY 181 AATCTCTAGCTTAAATAATGAAGCTCTGCTGTCTTTCTGATTCAGATGGCCATCTT 240  
Db 181 AATCTCTAGCTTAAATAATGAAGCTCTGCTGTCTTTCTGATTCAGATGGCCATCTT 240  
QY 241 GAGCTCTATCTTCCACATGTACAGCCACACACATCAGCTCCCTGCTATGAGGAGCAGCC 300  
Db 241 GAGCTCTATCTTCCACATGTACAGCCACACACATCAGCTCCCTGCTATGAGGAGCAGCC 300  
QY 301 CGAGCGCATGACAGCTGCTGTCTGTCTTCTTGCGGCTCTGGGCTCTTTTGTGGGCA 360  
Db 301 CGAGCGCATGACAGCTGCTGTCTGTCTTCTTGCGGCTCTGGGCTCTTTTGTGGGCA 360  
QY 361 GCTTTTGGGAGAGCACCAGATGTTTCTACCTGATGAGACCCCGCTGGAGCGTGTGAT 420  
Db 361 GCTTTTGGGAGAGCACCAGATGTTTCTACCTGATGAGACCCCGCTGGAGCGTGTGAT 420  
QY 421 GACCTTCAGAGAGCAGCCGCTGATGCTGCACATGGCTGTGGGAGATGTATACGGGC 480  
Db 421 GACCTTCAGAGAGCAGCCGCTGATGCTGCACATGGCTGTGGGAGATGTATACGGGC 480  
QY 481 CGTCTTCTGTGCACATGTAGACCGCTTTGATGCTTACATGGAACCTGTCCCCGAGACA 540  
Db 481 CGTCTTCTGTGCACATGTAGACCGCTTTGATGCTTACATGGAACCTGTCCCCGAGACA 540  
QY 541 GTCCAGCCTCTTTCAGTGGAGAAACAGCCGGGCCCTGTGTTCTGCACCTGCTGTGACAT 600  
Db 541 GTCCAGCCTCTTTCAGTGGAGAAACAGCCGGGCCCTGTGTTCTGCACCTGCTGTGACAT 600  
QY 601 CATCCCAAGATGAAATCATCCCCGGGGCTCAGTGCAGGCTCTCTGTGACATACAGCC 660  
Db 601 CATCCCAAGATGAAATCATCCCCGGGGCTCAGTGCAGGCTCTCTGTGACATACAGCC 660  
QY 661 CTTTGAAGTGTGAGAGAGGCTGCCGCTCTTACAGCCAGCTGTGCTCAAGAGGTGCG 720  
Db 661 CTTTGAAGTGTGAGAGAGGCTGCCGCTCTTACAGCCAGCTGTGCTCAAGAGGTGCG 720  
QY 721 CTTCTTAACTGAGCTCCTCTTACCCGCTGTGAAAGACCCCTCCCTCAACCTGCATAT 780  
Db 721 CTTCTTAACTGAGCTCCTCTTACCCGCTGTGAAAGACCCCTCCCTCAACCTGCATAT 780  
QY 781 CGTGCACCTGTGCGGGAGCCCGGGGCGGTTCGCTTCCCGAGAGCAGCAAGGAGGA 840  
Db 781 CGTGCACCTGTGCGGGAGCCCGGGGCGGTTCGCTTCCCGAGAGCAGCAAGGAGGA 840  
QY 841 TCTCATGATTGACAGTGCATGTGTGATGGGCGAGCATGAGCAGAAACTCAAGAGAGGA 900  
Db 841 TCTCATGATTGACAGTGCATGTGTGATGGGCGAGCATGAGCAGAAACTCAAGAGAGGA 900  
QY 901 CCAACCTTACTATGTGTGAGAGTTCATCTGCCAAAGCCAGCTGGAGATCTACAGACCAT 960  
Db 901 CCAACCTTACTATGTGTGAGAGTTCATCTGCCAAAGCCAGCTGGAGATCTACAGACCAT 960

QY 961 CCAGTCTTGGCCCAAGGCCCTGCAGAGCGTACCTGCTTGTGCGGTATGAGACCTGGC 1020  
Db 961 CCAGTCTTGGCCCAAGGCCCTGCAGAGCGTACCTGCTTGTGCGGTATGAGACCTGGC 1020  
QY 1021 TCGAGCCCTCTGTGGCCCAAGCTTCCGAATGTATGAAATTCGTGGGATTTGGAATTCCTGCC 1080  
Db 1021 TCGAGCCCTCTGTGGCCCAAGCTTCCGAATGTATGAAATTCGTGGGATTTGGAATTCCTGCC 1080  
QY 1081 CCATCTTCAGACCTGAGTGCATACATCAACCCGAGCAAGGGGATGGGATGACACGCTTT 1140  
Db 1081 CCATCTTCAGACCTGAGTGCATACATCAACCCGAGCAAGGGGATGGGATGACACGCTTT 1140  
QY 1141 CCACACAAATGCCAGGAGATGCCCTTATGTCTCCAGGCTTGGCGCTGCTTTGCCCTTA 1200  
Db 1141 CCACACAAATGCCAGGAGATGCCCTTATGTCTCCAGGCTTGGCGCTGCTTTGCCCTTA 1200  
QY 1201 TGAAGAGTTTTCGAGCTTCAGAAAGCCTGTGGCGATGGCATGAATTTGCTGGCTACCG 1260  
Db 1201 TGAAGAGTTTTCGAGCTTCAGAAAGCCTGTGGCGATGAATTTGCTGGCTACCG 1260  
QY 1261 CCACGTCAATCTGAACAAGAACAGAAACCTGTTGCTGATCTTCTGTCTACCTGGAC 1320  
Db 1261 CCACGTCAATCTGAACAAGAACAGAAACCTGTTGCTGATCTTCTGTCTACCTGGAC 1320  
QY 1321 TGTCCCTGAGCAAAATCCACTAAGAGGCTTGAGAAGGCTTTGCTGCCACCTGCTGTACGC 1380  
Db 1321 TGTCCCTGAGCAAAATCCACTAAGAGGCTTGAGAAGGCTTTGCTGCCACCTGCTGTACGC 1380  
QY 1381 TCAGTCACTTTCCTGAATGCTTCTGAGGCTTGCCTACATCTCTGATAGCTTAACTACATG 1440  
Db 1381 TCAGTCACTTTCCTGAATGCTTCTGAGGCTTGCCTACATCTCTGATAGCTTAACTACATG 1440  
QY 1441 TCTGTGGGATACACACTGTGAGTGTGTGCACACCTGCTCAAGCAGAGAGACTTTT 1500  
Db 1441 TCTGTGGGATACACACTGTGAGTGTGTGCACACCTGCTCAAGCAGAGAGACTTTT 1500  
QY 1501 GTGTCCATCTTGTGTCTGAAACAGACTGGGGAACCTTATGTGGAGCAGACATCCGC 1560  
Db 1501 GTGTCCATCTTGTGTCTGAAACAGACTGGGGAACCTTATGTGGAGCAGACATCCGC 1560  
QY 1561 CAGTGAACAGGATTTGCTCTTCTTCTTCTGTGATCTTCTGCTGTGGGAGACTTGA 1620  
Db 1561 CAGTGAACAGGATTTGCTCTTCTTCTTCTGTGATCTTCTGCTGTGGGAGACTTGA 1620  
QY 1621 AGACTTGTGGGCTGGAGGCTATTAGCAGACACAGATACAGTATGAGAAATGATCCATA 1680  
Db 1621 AGACTTGTGGGCTGGAGGCTATTAGCAGACACAGATACAGTATGAGAAATGATCCATA 1680  
QY 1681 ACCTCCCTGTCACATCTTGCCCAATGGGGAATGGAATCTTACCAAGAGGCTCACAGC 1740  
Db 1681 ACCTCCCTGTCACATCTTGCCCAATGGGGAATGGAATCTTACCAAGAGGCTCACAGC 1740  
QY 1741 ATTTTCCAGAGATGCAATTTCTAGCCCTTGAGGTTTCCCGATGGATTTCAAGAGAGAA 1800  
Db 1741 ATTTTCCAGAGATGCAATTTCTAGCCCTTGAGGTTTCCCGATGGATTTCAAGAGAGAA 1800  
QY 1801 GTGGGAACAAGGTGGATGCTTACTTATGAGGCTTACCATCAGACCTATCGCTAATCAGA 1860  
Db 1801 GTGGGAACAAGGTGGATGCTTACTTATGAGGCTTACCATCAGACCTATCGCTAATCAGA 1860  
QY 1861 AATATGAAACAANAATCTCGACAAAGAGCAAGCTTTTAAGTTTACAGAGGTGCTGGGC 1920  
Db 1861 AATATGAAACAANAATCTCGACAAAGAGCAAGCTTTTAAGTTTACAGAGGTGCTGGGC 1920  
QY 1921 TGCATTTGAATATCACTTCCCTCTGCATTTCCCATCAGATAGAGACTTTGACCTGTG 1980  
Db 1921 TGCATTTGAATATCACTTCCCTCTGCATTTCCCATCAGATAGAGACTTTGACCTGTG 1980  
QY 1981 AAGCTGCCATCTGTTAATCTAAATTCCTCAAAATAGAAAAAATTTTAAAAA 2032  
Db 1981 AAGCTGCCATCTGTTAATCTAAATTCCTCAAAATAGAAAAAATTTTAAAAA 2032



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:51:20 ; Search time 5293 Seconds

(without alignments) 11172.662 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032  
Sequence: 1 ggctcgagggccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_in:\*  
5: gb\_in:\*  
6: gb\_ov:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_iny:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2032	100.0	2032	6	AR203335	AR203335 Sequence
2	2032	100.0	2032	9	AF131235	AF131235 Homo sapi
3	1905.4	93.8	2011	9	AK026635	AK026635 Homo sapi
4	1857.8	91.4	1992	9	AF280088	AF280088 Homo sapi
5	1856.6	91.4	18328	9	AC010547	AC010547 Homo sapi
6	1261.6	62.1	1333	9	AF149783	AF149783 Homo sapi
7	782.4	38.5	2201	10	AF109155	AF109155 Mus muscu
8	741.6	36.5	1926	10	AF131236	AF131236 Mus muscu
9	506.4	24.9	517	6	AX381256	AX381256 Sequence
10	414.4	20.4	1647	6	AX327330	AX327330 Sequence
11	414.4	20.4	2544	9	AF219990	AF219990 Homo sapi
12	414.4	20.4	3786	9	AF280086	AF280086 Homo sapi
13	414.4	20.4	17503	9	AF219991	AF219991 Homo sapi
14	414.4	20.4	15737	9	AC009163	AC009163 Homo sapi
15	414.4	20.4	208185	2	AC009105	AC009105 Homo sapi
16	400	19.7	1462	9	AF176839	AF176839 Homo sapi
17	400	19.7	2170	9	AF176838	AF176838 Homo sapi
18	400	19.7	3278	9	AF246718	AF246718 Homo sapi
19	400	19.7	194832	9	AC025287	AC025287 Homo sapi
20	388.8	19.1	1740	10	AF176841	AF176841 Mus muscu
21	388.8	19.1	1989	10	AF176840	AF176840 Mus muscu
22	388.8	19.1	159170	2	AC115914	AC115914 Mus muscu
23	388.8	19.1	1237950	2	AC127315	AC127315 Mus muscu
24	368.8	18.1	133902	2	AC095664	AC095664 Rattus no
25	324	15.9	108261	2	AC130174	AC130174 Rattus no
26	287.4	14.1	389	6	AX381326	AX381326 Sequence
27	266	13.1	184551	2	AC117869	AC117869 Rattus no
28	146.4	7.2	6961	9	AB017915	AB017915 Homo sapi
29	146.4	7.2	178253	9	AC073370	AC073370 Homo sapi
30	146.4	7.2	196467	9	AC022392	AC022392 Homo sapi
31	143.2	7.0	2156	6	AR071396	AR071396 Sequence
32	143.2	7.0	2156	6	E14937	E14937 Human mRNA
33	143.2	7.0	2731	9	AB012192	AB012192 Homo sapi
34	135.2	6.7	200292	2	AC087442	AC087442 Homo sapi
35	133.6	6.6	1241	9	AF090137	AF090137 Homo sapi
36	133.6	6.6	1458	6	E16306	E16306 cDNA encodi
37	133.6	6.6	2190	6	AR211733	AR211733 Sequence
38	133.6	6.6	2190	9	HS05637	HS05637 Homo sapien
39	133.6	6.6	2415	9	AB003791	AB003791 Homo sapi
40	133.6	6.6	2453	9	BC028235	BC028235 Homo sapi
41	133.6	6.6	2734	9	BC022567	BC022567 Homo sapi
42	130	6.4	2354	5	CHKG6ST	CHKG6ST
43	130	6.4	2354	6	AR050577	AR050577 Sequence
44	130	6.4	2354	6	E12369	E12369 DNA encodin
45	123.8	6.1	2227	9	AF083066	AF083066 Homo sapi

#### ALIGNMENTS

RESULT 1  
AR203335  
LOCUS AR203335 2032 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 2 from patent US 6365365.  
ACCESSION AR203335  
VERSION AR203335.1 GI:21499698  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2032)  
AUTHORS Bistrup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.  
TITLE Method of determining whether an agent modulates glycosyl  
sulfoltransferase-3  
JOURNAL Patent: US 6365365-A 2 02-APR-2002;

Pred. No. is the number of results predicted by chance to have a